FROM META-ETHICS TO META-RATIONALITY

Antonio Gaitán Torres

Departamento de Filosofía I
Universidad de Granada

A María – y a nuestro Pablo

Agradecimientos

Índice

Introducción

Cualquier investigación está animada, desde sus inicios, por una serie de intuiciones o convicciones básicas. Estas intuiciones estructuran los diferentes estadios que componen la citada investigación, dotándola de unidad y coherencia. Como no podía ser de otro modo, algunas de estas intuiciones son claramente erróneas y una buena investigación debería motivar su rechazo. Otras intuiciones, en cambio, no se erigen como completamente desencaminadas; incluso cuando hemos avanzado un buen trecho apreciamos su creciente plausibilidad,

Todo esto viene al caso porque cuando comencé a trabajar en esta tesis una intuición me parecía tremendamente plausible. Por aquel entonces creía que cuestiones centrales para la *Filosofía de la Mente* - entendida de modo amplio, englobando lo que denominamos *Filosofía de la Acción* y lo que solemos entender como *Teoría de la Racionalidad* – eran esenciales para entender la naturaleza de nuestros juicios morales. Después de escribir este trabajo de investigación mi grado de compromiso con esta intuición no ha variado de modo significativo. En sentido estricto sigo creyendo que podemos y debemos hacer *Meta-ética* de modo oblicuo. Podemos acometer aquellas cuestiones que definen la *Meta-ética* como disciplina filosófica a través de una investigación centrada en ciertos temas que son tratados dentro de lo que arriba he caracterizado como *Filosofía de la Mente*. En este trabajo esbozo una posible ruta para desarrollar esta intuición. Lo que presento a continuación es la estructura de mi tesis en relación con este objetivo primario.

1. *Meta-ética*, práctica moral y desacuerdo

Según una visión bastante extendida, hacemos *Meta-ética* cuando nos preguntamos por una serie de cuestiones de segundo nivel centradas en nuestras opiniones morales. Cuestiones acerca del significado de esas opiniones, su estatuto ontológico y el tipo de evidencia que podríamos aducir para justificarlas configurarían, según esta concepción estándar, la perspectiva propia de la *Meta-ética*, su contenido. La amplia aceptación que goza esta caracterización del objeto de la *Meta-ética* no debería ocultar, sin embargo, un problema

bastante evidente. Por decirlo de modo breve: una vez que asumimos que la *Meta-ética* trata de formular ciertas cuestiones de segundo nivel sobre nuestras opiniones morales, ¿a qué tipo de asidero podemos recurrir para identificar las opiniones que constituyen el objeto de la *Meta-ética*? ¿Debemos asumir que son entidades o actos aisladas o, por el contrario, debemos incluirlas dentro de prácticas más complejas? ¿Cómo separamos una opinión genuinamente moral de una que no lo es? ¿Cómo caracterizamos, en suma, el foco sobre el cual la *Meta-ética* teoriza?

En el primer capítulo de mi tesis defiendo que un modo razonable para fijar el foco de interés de la *Meta-ética* atendería a nuestras intuiciones sobre nuestras opiniones morales. En ese capítulo asumo que un buen modo de saber qué opiniones caen dentro del dominio de la *Meta-ética* pasaría por explicitar aquellas regularidades, disposiciones conductuales y expectativas en las que normalmente encuadramos nuestro discurso moral. Sólo aquella perspectiva evaluativa que satisficiera de modo ajustado esa descripción podría conformar el objeto de estudio de la *Meta-ética*. Al hacer *Meta-ética,* por tanto, estamos interesados en explicar aquella perspectiva evaluativa que instancia nuestras intuiciones sobre nuestra práctica moral. Explicar esas intuiciones sirviéndonos del aparataje conceptual ofrecido por la *Semántica*, la *Epistemología* o la *Ontología* equivale a *acomodar* nuestra perspectiva moral. La *acomodación* de esa perspectiva moral constituye el ideal metodológico básico que ánima a la *Meta-ética*.

Clarificar este ideal metodológico anuncia una posibilidad incómoda. Si al hacer *Meta-ética* acomodamos nuestra perspectiva moral, ¿no podría darse una situación en la que dos teorías de segundo nivel totalmente incompatibles entre sí (dos teorías que ofreciesen explicaciones opuestas sobre de la *ontología*, la *semántica* y la *epistemología* de nuestros juicios morales) explicasen o acomodasen de modo igualmente plausible las intuiciones que conforman nuestra perspectiva moral?

En el capítulo segundo asumo que está posibilidad se ilustra en el panorama meta-ético actual, donde dos teorías meta-éticas (*cognitivismo* y *no cognitivismo*) ofrecen dos explicaciones igualmente consistentes de nuestras intuiciones básicas - sin compartir supuestos centrales a nivel filosófico. En este capítulo analizo las consecuencias metodológicas que podríamos extraer de esta situación de indefinición. Allí señalo cuatro posibles alternativas para hacer frente al *impasse* en el que se haya sumida la *Meta-ética* (A y

B se interpretan, en lo que sigue, como dos teorías meta-éticas que asumen un dominio común de intuiciones sobre nuestros juicios morales):

- (i) Podríamos argumentar que el *impasse* entre A y B, aunque conceptualmente posible, no es real. Podríamos aducir, por ejemplo, que cierta intuición central para entender la naturaleza de nuestra perspectiva moral es acomodada por A pero no por B (o no por B con la misma suficiencia que A).

- (ii) Podríamos asumir que el *impasse* entre A y B es real, contrarrestándolo a través de la inclusión de otros criterios adicionales, unos que vendrían a completar el ideal de *acomodación* esbozado arriba. Podríamos señalar, por ejemplo, que B es una mejor elección teórica que A porque ésta acarrea una serie de consecuencias prácticas o existenciales desastrosas – aunque sus postulados sean filosóficamente respetables.

- (iii) Podríamos aceptar la realidad del *impasse* entre A y B extrayendo como consecuencia una visión escéptica acerca de las ambiciones metodológicas de A y B y de la *Meta-ética* en general.

- (iv) Finalmente, podríamos admitir el *impasse* entre A y B sin comprometernos ni con un escepticismo metodológico general (caso de (c)) ni con ciertas consideraciones sustantivas encaminadas a contrarrestar los aspectos negativos del *impasse* (caso de b).

En mi trabajo defiendo - contra lo apuntado en (i) - que el *impasse* en *Meta-ética* es real. De entre aquellas opciones que aceptan la realidad del *impasse* abogo por una versión de (iv). Si bien acepto que actualmente tanto los cognitivistas como los no cognitivistas son capaces de acomodar nuestras intuiciones morales con similar grado de precisión, creo que la situación de indefinición podría minimizarse si atendiéramos a nuestros juicios morales desde una perspectiva diferente. Esta perspectiva trataría de acentuar la centralidad de nuestras *razones* a la hora de comprender los fundamentos de nuestra práctica evaluativa. Si atendiésemos a los diferentes usos asociados a nuestro concepto de *razón* (usos ligados a la explicación y justificación de nuestras acciones) podríamos obtener una perspectiva analítica

adicional desde la cual facilitar una posible solución a la situación de indefinición teórica que acabo de esbozar.

2. *Meta-ética*, razones motivacionales y razones normativas

La descripción y evaluación de dos posibles versiones de (iv) ocupa la parte central de mi tesis. En este estadio intermedio presento dos argumentos que se han venido utilizado para sustentar una determinada tesis meta-ética partiendo del análisis de nuestro concepto de *razón*. Estos argumentos tienen, como se verá en un momento, bastantes cosas en común.

Meta-ética y motivación

En el capítulo 3 presento un argumento que parte del concepto de *razón* en sentido *explicativo* y que se sirve de ciertas intuiciones que remarcan la fuerza motivacional de nuestras opiniones morales. Si este argumento fuese válido, una teoría moral de tipo no cognitivista se vería tremendamente reforzada. El argumento que me interesa se estructura del modo siguiente:

(1) Necesariamente, si A profiere en t_1 un juicio con contenido moral – 'ϕ es correcto' - entonces A intentará hacer ϕ en t_2 (*Internismo*)

(2) A haría ϕ intencionalmente si y sólo si ϕ pudiera asociarse con una pro-actitud de A hacia ϕ (*Teoría humeana de la motivación*)

En consecuencia, si (1) y (2) son verdaderas, la proferencia de un juicio con contenido moral – 'ϕ es correcto' – expresaría una pro-actitud hacia ϕ - un *deseo* de que ϕ sea el caso, una *prescripción* centrada en ϕ, o un *compromiso* con un sistema de normas que prescribe ϕ en ese contexto concreto.

A lo largo del capítulo 3 paso revista a algunas de las críticas que este argumento ha recibido. Estas críticas descansan, como cualquier crítica realizada sobre cualquier argumento, en un proceso de *triangulación*. Podríamos negar la conclusion del citado argumento si

remplazásemos o criticásemos cualquiera de las asunciones que conforman sus premisas. No obstante, aunque *triangular* podría ser una opción natural llegados este punto, la opción que se favorece en este trabajo para rechazar el argumento motivacional, cualquier argumento motivacional de hecho, es ligeramente diferente. En el capítulo 3 sugiero algo bastante simple: si probásemos que alguna de las asunciones generales que animan la estructura argumentativa esbozada arriba no puede formularse de modo preciso (porque está sujeta a un desacuerdo esencial, porque es esencialmente ambigua, o porque no puede formularse de modo claro) entonces podríamos colegir que cualquier argumento que apela a esa asunción tiene un dudoso valor. Por tanto no necesitaríamos probar la falsedad de una determinada premisa para rechazar cualquier variante del argumento esbozado arriba. Con establecer que una determinada asunción está sujeta a un desacuerdo genuino tendríamos una base suficiente para echazar cualquier argumento que hace uso de esa asunción.

La asunción que he elegido para ilustrar esta estrategia es aquella que trata de establecer el tipo de vínculo que existe entre nuestros juicios morales y nuestra voluntad. El desacuerdo en torno a esta asunción es muy profundo. Eso hace que cualquier argumento que apela a la asunción motivacional tiene pocos visos de establecer una conclusión sólida a favor de una determinada teoría meta-ética. Puesto que ninguna de las interpretaciones mayoritariamente asumidas sobre la fuerza motivacional de nuestros juicios morales (*internismo* o *externismo*) puede sustentar un argumento concluyente a favor de una determinada concepción meta-ética, debemos asumir que ningún argumento en clave motivacional puede cancelar el *impasse* que atenaza a la *Meta-ética*.

Meta-ética y justificación

La búsqueda de argumentos concluyentes a favor de una u otra concepción meta-ética no se detiene en el anterior intento. Además de aquellos contextos en los que *explicamos* lo que A hizo apelando a sus motivos, ciertas prácticas discursivas se estructuran en torno a la *justificación* de las acciones de A. En los contextos en que esas prácticas se insertan aludimos a las *razones* que A tiene en un sentido *normativo*. Una *razón* en sentido normativo suele referir a un hecho que favorece o justifica una determinada conducta o actitud. Si A tiene una razón para hacer ϕ, seguramente podemos decir algo a favor de ϕ. Esta dimensión normativa es inherente a nuestro concepto de *razón y* como tal es el foco de otro argumento clave para la

Meta-ética. Si este argumento fuese exitoso, una ruta para evitar el *impasse* metodológico se nos abriría de modo claro. El argumento se puede esquematizar del siguiente modo:

(i) Necesariamente, si A está moralmente obligado a hacer ϕ entonces A tiene una buena razón para hacer ϕ (*Racionalismo*)

(ii) Necesariamente, si A tiene una buena razón para hacer ϕ entonces A puede estar motivado a hacer ϕ bajo ciertas condiciones (*Internismo normativo*)

(iii) A está motivado a hacer ϕ si y sólo si ϕ-ing puede ser asociada con una pro-actitud de A hacia ϕ (*Teoría humeana de la motivación*)

En consecuencia, si (i)-(iii) son verdaderas, A esta moralmente obligado a hacer ϕ sólo si ϕ-ing puede asociarse con una pro-actitud de A hacia ϕ

Este argumento sistematiza una serie de intuiciones o asunciones generales sobre nuestra perspectiva moral. Por ejemplo, es razonable suponer que nuestra perspectiva *moral* no difiere de aquella perspectiva que ocupamos cuando tratamos de responder a cualquier *razón*, es decir, a cualquier obligación fundamentada en el hecho de que cierto rasgo favorece una determinada respuesta por nuestra parte. Igualmente, parece plausible comprometerse con una visión muy extendida sobre el estatuto de nuestras razones normativas según la cual nuestras razones se fundamentan necesariamente en nuestros motivos. Evidentemente, una vez que el *racionalismo moral* y el *internismo normativo* son asumidos, una conclusión parece imponerse de modo natural: nuestras obligaciones morales - aquellas que explicitan la relevancia de cierto tipo de razones a la hora de justificar nuestra conducta - apelan en última instancia a nuestros motivos, a nuestros deseos o nuestras preferencias debidamente corregidas. Por tanto, en la medida en que la *normatividad* de nuestras obligaciones morales (el modo en que éstas guían nuestra conducta más allá de la coacción o la fuerza física) debe explicarse haciendo referencia a nuestros motivos, es necesario asumir que nuestras opiniones morales expresan, en última instancia, un subconjunto (debidamente informado y corregido) dentro del conjunto conformado por nuestras preferencias o deseos.

Mi estrategia en relación con este argumento es idéntica a la que apliqué anteriormente en

el contexto motivacional. De entrada, no afirmo que una determinada interpretación de las asunciones esbozadas arriba sea verdadera o falsa. Mi objetivo durante los capítulos 5 y 6 es, por contra, persuadir al lector de que ninguna interpretación del marco general que sustenta el citado argumento está libre de polémica. Y no lo está por el mismo tipo de consideración estructural que aduje antes para rechazar los argumentos motivacionales en bloque. De la misma manera que antes era improbable lograr un acuerdo en torno a una de las intuiciones que componen la estructura de cualquier argumento motivacional (¿debe nuestra motivación moral ser entendida en clave interna o externa?), cuando atendemos a una de las intuiciones que subyacen al argumento que nos ocupa el acuerdo en torno a su contenido y formulación no resulta menos improbable. Me refiero, por supuesto, a la intuición centrada en el estatuto de nuestras razones normativas. Esta intuición se formula arriba como premisa ii. En el capítulo 6 defiendo que en la medida en que no podemos alcanzar un consenso en torno al estatuto de nuestras razones normativas, cualquier argumento que pretenda cancelar el *impasse* meta-ético apelando a esa intuición debe rechazarse. Es poco probable, sugiero, que podamos reiniciar la *Meta-ética* a través de este tipo de argumentos.

Para fundamentar esta conclusión negativa empiezo proponiendo una interpretación corregida de la premisa problemática. En este caso propongo una interpretación generosa de la premisa *internista* a nivel normativo. Una vez asumida esa interpretación (5.4 y 5.5), critico la formulación propuesta y las intuiciones que la animan (6.2 y 6.3). Finalmente, el hecho de que tampoco la imagen negativa que emerge de esta crítica (*externalismo normativo*) sea aceptable me empuja a afirmar que la intuición general a la que estas dos tesis meta-normativas tratan de dar forma está sujeta a un profundo desacuerdo (6.4). Ningún argumento basado en una intuición sobre el estatuto de nuestras razones normativas tiene visos de erigirse como un argumento concluyente a favor de una determinada tesis meta-ética. Si queremos reiniciar la *Meta-ética* a través de un enfoque oblicuo - uno centrado en el concepto de *razón* - deberemos mirar en otra dirección.

3. Meta-ética y Meta-racionalidad

Hasta ahora he sugerido que la metodología asumida por gran parte de aquellos que hacen *Meta-ética* hace concebible, al menos en principio, cierta situación de indefinición teórica. En esta situación, dos teorías podrían acomodar de modo igualmente ajustado una serie de intuiciones sobre nuestra práctica moral. Esta paridad teórica haría complicado decidir qué imagen filosófica general puede dotar de sentido a esa práctica. Sería complicado decidir qué ontología, qué teoría del significado, o qué epistemología podría explicar los contornos de nuestra experiencia moral. No tendríamos claro, por decirlo de algún modo, qué tipos de hechos sustentan nuestras opiniones morales. Tampoco qué función anima nuestro discurso moral y cómo esa función se relaciona con el tipo de justificación que deberíamos presuponer en aquellos casos en los que decimos saber que algo es correcto o incorrecto.

Para paliar esa sensación de abandono teórico recurro de nuevo al concepto de *razón*. En lugar de apelar a ciertas tesis sustantivas sobre nuestras *razones*, me centro en aquellas capacidades psicológicas que ejemplificamos al ser racionales, al responder a cierta variedad de razones ligadas a nuestro buen funcionamiento mental. En la parte final de mi tesis defiendo que si nos centrásemos en este fenómeno general (*reasons responsiveness*) podríamos establecer una buena piedra de toque para cancelar la situación de indefinición teórica dentro de la *Meta-ética*. Para contextualizar esta intuición metodológica propongo un marco general desde el cual reiniciar la *Meta-ética*. Éste asume tres puntos básicos.

(a) Cuando hacemos *Meta-ética* tratamos de dilucidar el tipo de estado psicológico que expresamos a través de nuestras opiniones morales - *Semántica en clave psicológica*

(b) Aquellos rasgos que son moralmente significativos (*que ella necesita ayuda, que sería incorrecto coger ese dinero*) facilitan *razones* para justificar una determinada acción - *Racionalismo moral*

(c) *Meta-racionalidad* (M) está interesada en dos cuestiones básicas: (i) qué concepto es más básico en el ámbito normativo y (ii) qué tipo de estado mental expresamos al responder a nuestras obligaciones de *racionalidad*.

(a) y (b) parecen bastante plausibles. Al menos tanto como dotar de contenido a (c.i) al suponer que nuestro concepto de *razón* es básico en sentido normativo, es decir, que la normatividad de otros conceptos y otras obligaciones puede explicarse apelando al concepto de *razón*. Pero en la medida en que (a) y (b) resultan aceptables y (c.i) se interpreta como he sugerido, una sugerencia metodológica se impone: podríamos ofrecer una respuesta a ciertas cuestiones *meta-éticas* – cuestiones centradas en el estatuto psicológico de nuestras opiniones morales - analizando el tipo de estados mentales que posibilitan nuestra capacidad para responder a ciertas obligaciones (obligaciones de racionalidad) susceptibles de ser justificadas en clave de *razones*. La tercera parte de mi tesis evalúa la plausibilidad de este enfoque general (capítulo 7).

Lo que aquí vengo denominando *Meta-racionalidad* se ha tratado de sustantivar de tres modos distintos, todos ellos centrados en (c.ii):

- (1) Se ha centrado en ciertas situaciones en las que evaluamos la racionalidad de un determinado sistema intencional para determinar qué tipo de estados mentales subyacen a nuestras adscripciones de racionalidad

- (2) Se ha ocupado de aquellas situaciones en las que dos agentes, A y B, ejemplifican un *genuino desacuerdo normativo* (GD) sobre *p*, es decir, un desacuerdo sobre cuándo deberíamos aseverar *que p* o planear *que p* fuese el caso. A y B ejemplifican un GD cuando el desacuerdo no puede resolverse apelando a ulterior evidencia (porque A y B conocen todos los hechos relevantes) y cuando el desacuerdo no descansa en ningún tipo de error deliberativo (porque A y B son óptimamente coherentes y racionales). Casos de desacuerdo entre iguales - *peer disagrement* - ilustran lo que trato de esbozar aquí. Una vez asumida la plausibilidad de un genuino desacuerdo normativo de tipo moral - GD-*moral* - ¿por qué no comparar las reacciones o disposiciones de A y B en un GD-*moral* con aquellas que tendrían en situaciones donde el desacuerdo es factual - GD-*no moral*? Pero al comparar nuestras reacciones o disposiciones ante estos dos tipos de escenarios, una hipótesis resulta bastante plausible: Si las disposiciones que ejemplificamos en aquellos casos de GD-*moral* varían de modo significativo en relación con aquellas que mostramos en casos de GD-*no moral* podría concluirse que no aseveramos nada en aquellos casos en los que estamos envueltos en GD-*morales*.

Porque, ¿cómo presuponer que aseveramos algo en sentido estricto en un GD-*moral* si no instanciamos algunas de las disposiciones típicas de nuestros usos asertivos, es decir, de aquellos usos sobre los que versan nuestros desacuerdos factuales?

- (3) Finalmente, se ha preguntado por el andamiaje psicológico que subyace a nuestros conceptos normativos de *razón* o *racionalidad* centrándonos en la relación que estos conceptos guardan con nuestras *decisiones*. Según esta tercera opción, la *Meta-racionalidad* analizaría el tipo de estado mental que expresamos cuando intentamos hacer *p* después de atribuir a *p* (entre otras opciones disponibles) la propiedad de *ser favorecido por ciertas razones* o, como Gibbard señala, la propiedad de *ser lo que debe hacerse*.

En el capítulo 7 me ocupo de las dos primeras opciones. Allí presento y critico dos argumentos concebidos a partir de aquellas posibilidades apuntadas por (1) y (2). Ambos pretenden facilitar dos rutas directas para sustentar una determinada concepción meta-ética, a partir de ciertos hallazgos sobre el estatuto psicológico de ciertos juicios normativos no específicamente morales. En la medida en que (3) requiere un desarrollo más detallado que el que posibilitaría un argumento directo, (3) queda inexplorada en este trabajo a este nivel.

En el capítulo 8 expongo una propuesta alternativa sobre el alcance de lo que vengo denominado *Meta-racionalidad*. Mi propuesta abarca los dos ámbitos e los que la Meta-racionalidad puede desarrollarse, (c.i) y (c.ii). Ésta es la parte más personal de la investigación. Trato aquí de responder a la siguiente cuestión: ¿Por qué no considerar aquellas situaciones en las que cumplimos con ciertos principios básicos de racionalidad como la piedra de toque para entender el tipo de estado mental que ejemplificamos en aquellos casos en los que *respondemos* a ciertas obligaciones *morales*? Dicho de otro modo: ¿Por qué no analizar aquellas estructuras psicológicas que posibilitan nuestro seguimiento de ciertos principios básicos de racionalidad para establecer el tipo de estado mental que expresamos al responder a ciertas demandas *morales*?

Acotar esta intuición metodológica general es una tarea compleja. Gran parte del capítulo 8 se dedica a precisar el modo en que cierto análisis de la *normatividad* ligada a nuestros principios básicos de racionalidad podría sustantivar esa intuición. Lo primero que hago en ese capítulo es explicitar una hipótesis muy extendida. Según ésta, la *normatividad* ligada a

nuestras obligaciones de racionalidad se podría explicar a partir de (i) aquellos rasgos valiosos a los que respondemos al cumplir con nuestras obligaciones de racionalidad y de (ii) aquellas reacciones negativas que otros agentes expresan cuando ignoramos esos rasgos valiosos, es decir, cuando incumplimos una determinada regla de racionalidad. Nuestro *debe* de *racionalidad* no es más que un indicador que rastrea una variedad determinada de *razones* ligadas a cierto tipo de *valores* (*coherencia*, *información*, *verdad*). Tal y como presento las cosas durante la parte final de mi investigación, si el siguiente principio meta-normativo fuese verdadero

> **(NR)** Necesariamente, si A está *obligado* a formar un determinado estado mental
> (ϕ) entonces A tiene una *razón* (concluyente) a favor de ϕ

entonces tendríamos una base suficiente para variar el foco de la *Meta-ética* hacia cuestiones de *Meta-racionalidad*.

Lo que acabo de sugerir requiere un argumento detallado. Facilito este argumento al inicio del capítulo 8 (8.1). Allí explico por qué la normatividad ligada a nuestras obligaciones racionales puede explicarse (y justificarse) en virtud de cómo esas obligaciones rastrean ciertas rasgos valiosos, ciertas razones. Adicionalmente establezco que las razones a las que respondemos al comportarnos *moralmente* son muy similares (tanto en su estructura y como en su categoricidad) a aquellas razones facilitadas por nuestras reglas de *racionalidad*. Ambas tesis meta-normativas sustentan NR. Y NR constituye el componente central de la tesis metodológica por la que abogo en este trabajo. A saber: que podemos aprender algo sobre el andamiaje psicológico que subyace a nuestra conducta y evaluación *moral* si analizamos nuestras respuestas ante ciertas reglas de *racionalidad*. En la medida en que la plausibilidad de esta tesis metodológica depende, en último extremo, de la plausibilidad de las dos asunciones meta-normativas que acabo de esbozar, dos cuestiones deben ser aclaradas.

De un lado, debemos preguntarnos si es verdad que la normatividad o la fuerza de nuestros principios de racionalidad - por ejemplo, el *principio de racionalidad instrumental* (PRI) - está necesariamente asociada con ciertos valores ¿Es verdad que *siempre* podemos explicar la normatividad ligada a PRI apelando a razones, es decir, a cierto tipo de valores que respetaríamos o fomentaríamos si formásemos nuestras intenciones de acuerdo con lo que PRI

demanda? Respondo a esta cuestión analizando una reciente discusión entre Niko Kolodny y John Broome (8.2).

De otro lado, ¿resulta plausible identificar las razones facilitadas por nuestras obligaciones *racionales* y por nuestras obligaciones *morales* únicamente a partir de un mismo tipo de *categoricidad*? Dicho de otro modo: aun si el paralelismo en torno a la *categoricidad* parece sugerir que la *Meta-ética* y la *Meta-racionalidad* se ocupan del mismo tipo de hechos normativos, ¿es suficiente con esta analogía para asegurar una reducción metodológica como la que este trabajo defiende? ¿No existen, después de todo, algunas diferencias importantes entre nuestras reglas morales y nuestras reglas de racionalidad – por ejemplo, que las segundas se aplican a estados mentales que a veces caen fuera de nuestro control voluntario (*creencias*, *deseos*, etc) mientras que las primeras regulan básicamente acciones?

Una vez asegurados estos dos puntos (8.3), la última parte de este trabajo trata de elaborar una imagen del tipo de psicología ejemplificada por aquellos *deliberadores* que son capaces de ser *racionales*, es decir, capaces de responder a aquellas razones o valores rastreados por nuestros principios básicos de racionalidad. Al final de mi investigación trato de responder a una única cuestión relacionada con las capacidades psicológicas que nos permiten cumplir con nuestras reglas de racionalidad a través de un proceso deliberativo (8.4). Supongamos que respondemos a nuestras reglas de racionalidad formando un determinado tipo de creencias: ¿cómo deberíamos interpretar los hallazgos psicológicos apuntados desde *Meta-racionalidad* a la hora de solucionar el supuesto *impasse* en *Meta-ética*? Aquí mi foco de interés se traslada a cuestiones que tienen que ver (a) con el estatuto y función de nuestras representaciones - así como con aquellos actos lingüísticos a través de los que expresamos esas representaciones – y con (b) el modo correcto de modelar la deliberación teniendo en cuenta la perspectiva del agente que delibera. En cierto sentido, trato de reubicar la importancia de esas cuestiones para responder en positivo a la situación de indefinición metodológica en la *Meta-ética* se instala incluso cuando es construida como *Meta-racionalidad*.

Chapter 1.

The complexity of our moral stance

1. Introduction

2. An initial characterization of *Meta-ethics*

3. *Accomodation* as a methodological ideal

4. The importance of basic platitudes about our moral stance

4.1 Formal platitudes about our moral stance

4.2 Substantives platitudes about moral stance

5. Looking forward

Chapter 1

1. Introduction

The present work is devoted to argue for a single although somehow complex claim. I will defend in this study that we could achieve a better stance to answer some meta-ethical questions if we developed a sound second-order account of the normative phenomenon exemplified by our responsiveness to reasons. The basic claim I want to argue for could be put in the following way:

> **[M]** If we attained a second-order understanding of our general responsiveness to reasons, then we would get a better understanding of the second-order psychological nature of our moral judgments

Some people would require a further thesis to secure the overall plausibility of M, even if a methodological framework for M were available. This should establish that moral judgments imply judgments about the reasons we have. In philosophical parlance this intuition has been usually called *rationalism*. Any version of *rationalism* will assume the following general claim:

> **[R]** Any moral judgment expressed by A is conceptually connected with a judgment about a certain subset of normative reasons

In sum, if *rationalism* - as it is encapsulated in R - were true, then we could do *Meta-ethics* by enquiring for the second-order nature of those psychological processes where we are responding to reasons - as M is defending. *Meta-ethics* could be reduced to *Meta-rationality*

By locating my investigation in the domain of *Meta-rationality* (M) and by further endorsing a sort of substantive point about the status of our moral commitments (R), I am presenting my topic – a meta-methodological sort of topic. In order to establish the priority of a second-level approach to morality based on the related concepts of *reasons* and *rationality* I will be responding to three basic questions:

(a) What is *Meta-ethics?* Or more particularly, what is the proper focus of *Meta-ethics*? Why am I presupposing that it would be better understood as concerned with the *psychological* content of our moral opinions rather than with any *ontological*, *epistemological* or *semantic* sort of inquiry? Why might we assume that *Meta-ethics* (once it is understood along the psychological lines noted above) would fail to offer a direct response to the question about the nature of our moral judgments or, more generally, of our moral opinions?

(b) What is *Meta-rationality?* Or more particularly, why am I presupposing that the proper focus of *Meta-rationality* is also a psychological-oriented sort of question, i.e. one focused on the psychological structures underlying our responsiveness to reasons? Why not to look for a more ontological, semantic or epistemic inquiry about reasons and rationality?

(c) Finally, what kind of conceptual link, would allow us to affirm that it could be possible to reduce *Meta-ethics* to *Meta-rationality*? Why would we achieve an understanding of what psychological state is being expressed by a moral claim by previously knowing what sort of mental state is enabling our capacity to recognize reasons and our disposition to be rational?

2. An initial characterization of *Meta-ethics*

As I said before, a great part of this investigation will deal with the plausibility of a certain non-standard image of *Meta-ethics*, one where issues of *Meta-rationality* are located in the foreground of the discussion. Evidently, in order to grasp such meta-methodological claim we should clarify what *Meta-ethics* is. In the current chapter I will be concerned with a partial aspect of this preparatory task. I will be interested in getting a minimal understanding of what *Meta-ethics* is about. To put it in another words, I will be interested in achieving a clear image

of the nature of our first-order moral domain, i.e. the domain of opinions being the focus of *Meta-ethics*.

But what do we understand by *Meta-ethics*? It is traditionally defended that by *Meta-ethics* we are simply characterizing a second-level perspective toward our moral opinions, a perspective concerned basically with questions about the meaning of the terms contained in our evaluative opinions, the ontological status associated with such terms and, finally, the sort of evidence required to justify those opinions endorsed from a genuine moral stance. A standard characterization of *Meta-ethics* should assume that it is mainly concerned with question about the *semantic*, *ontology*, and *epistemology* of our moral opinions. Let us consider, for instance, the definition offered by David McNaughton:

> "(…) there are questions about the nature of and status of our moral thought: Are there any moral truths? It is possible to show that one moral view is better than another? (…) Thinking about the status of moral thought is sometimes called *meta-ethics* to distinguish this last approach from practical ethics and the construction of moral theories" McNaughton, D. 1988. p. 15-16

Or a recent statement by Horgan and Timmons:

> "Metaethics is primarily concerned with semantics, metaphysics, and epistemology of moral thought and discourse" Horgan, T. Timmons, M. 2006a. p. 220

Or, finally, the characterization endorsed by Michael Smith in *The Moral Problem*:

> "(…) in meta-ethics the focus is not the substance of morality – not on specific recommendations about how to act or what to prefer or the principles that guide such recommendations – but is rather on a range of interrelated semantic, metaphysical, psychological, and epistemological questions about the practice of making moral judgments themselves" Smith, M. 2005. p. 6

Clearly, all these statements are agreeing, at least, on two basic points.

Firstly, they are agreeing on the centrality of certain second-order concerns in defining the shape and content of *Meta-ethics*. In a certain way, and by restricting ourselves to the quotes above, it is easy to note how, while Smith and McNaughton both acknowledge the centrality of a psychological-oriented stance focused on our moral opinions, Timmons and Horgan assume, on the contrary, a view in which meta-ethics is primarily concerned with a set of semantic, ontological and epistemological questions.

Secondly, the quotes above implictly acknowledge the centrality of our moral practices, 'the practice of making moral judgments' in Smith's own words, to define the domain that concerns *Meta-ethics*. They are all agreeing on a simple intuition: A *meta-ethical* question is a question *about our moral stance*, i.e. about the peculiar *perspective* we occupy when we utter a moral sentence.

But if the just noted distinction is sound we should divide the definitional task around *Meta-ethics* in two basic issues. *One the one hand*, we should take into account the nature of the discursive *domain* that is the first-order focus of *Meta-ethics*. *On the other hand*, we should determine the *conceptual tools* associated with the second-order domain from where we are going to offer a second-order explanation of our moral realm:
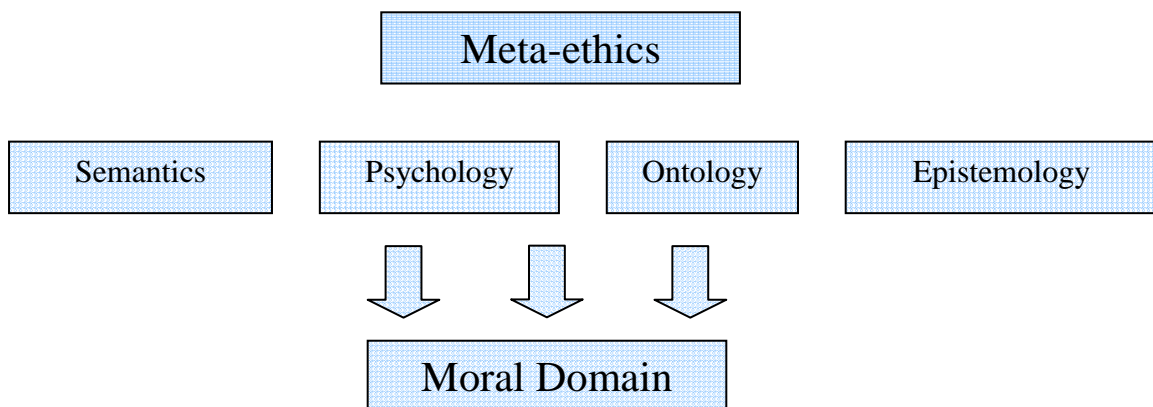
Accordingly, the questions that further follow about the status of *Meta-ethics* could be divided into two basic types of questions:

(a) Questions about the nature of the discursive and practical realm that is the *focus* of *Meta-ethics*

(b) Questions about the *theoretical components* required to offer an overall and plausible second-order approach on this first-order realm

In this chapter I will deal with (a) by offering an account of the main features associated with our moral domain.

3. *Accommodation* as a methodological ideal

If the previous sketch is minimally accurate, in order to properly explain what *Meta-ethics* is about we should first isolate our *intuitions* or *platitudes* about the stance we occupy when we utter a moral sentence. To determine our intuitions about our *moral stance* is important because *Meta-ethics* tries to answer some fundamental questions about the status and objectivity of our moral opinions through the *accommodation* of a set of intuitions focused on what we are doing when we evaluate a certain item in moral terms. In general, to *accommodate* a certain practice is going to be understood hereafter along *explanatory* lines. We *accommodate* the set of platitudes that define a certain practice when we locate such platitudes into the explanatory mould offered by a set of second-order disciplines. In the case at hand, we *accommodate* the set of platitudes that shape the image we have about our *moral stance* when we make sense of such platitudes from the explanatory standpoint offered by *Semantics*, *Epistemology,* and *Ontology*. By doing so we offer a *proper* account of the status of our moral opinions [Smith, M. 1996. Chapter 1. Timmons, M. 1999. Chapter 1. Railton, P. 1996. p. 59-66. 2006b. p. 202-205. Gibbard, A. 2003. p. 186-187 and 2007 p. 694. Cuneo, T. 2007. p. 32][1]

---

[1] Evidently, I am assuming a meta-philosophical thesis here about what a philosophical account of X would be. My commitment at this level is close to Pettit's: "We always do philosophy in dialogue with positions that already have a hold on us. Philosophy, as we might put it, is an attempt to come to terms with those opinions, endorsing them if they prove worthy of reasoned endorsement and seeking to liberate ourselves of them otherwise (…) Philosophy will be conducted then as a field where rivals forces pull against each other (…) fidelity to the manifest image of how things are, on the one side; fidelity to the intellectual image of how things

The notion of *accommodation* is central for *Meta-ethics* nowdays [Timmons, M. 1999. p. 11-12]. In some way, a great number of meta-ethicists would accept that the merit of their second-order explanations of the status of our moral opinions depends on the number of platitudes about our moral stance being *accomodated* by such explanations. They would be happy to accept that the accuracy of a given meta-ethical theory is proportional, so to speak, to the way in which such theory is making sense of our core intuitions about our moral perspective. *Accomodation* has emerged, in some sense, as a paramount methodological ideal for those who are doing *Meta-ethics.*

Surely we need to slow down here. Let assume that certain platitudes about our moral stance are very important from a metahodological standpoint. How do we know them? Why are we so sure about them? Moreover, do we really have intuitions about our moral stance? To minimize this initial concern we need to remind an obvious fact. It is pretty clear that we have intuitions about the evaluative status of certain items. We usually claim to have strong intuitions about whether certain action-types are *right*, about whether certain features of an object or certain experiences are *good*, about whether a certain character trait is *praiseworthy* or *blameworthy*, or about whether a given institution or social agreement is *fair* or *unfair.* These intuitions help us to theorize about our moral practice. They offer an observational basis to systematize the basic shape of our moral thought. And this systematization is the focus of what is commonly referred to as *Normative Ethical Theory* [Smith, M. 1994. p. 1-2. Timmons, M. 1999. p. 9. Kamm, F. M. 2006. Chapter 1]. If I am right, however, besides of these intuitions about the deontic and evaluative status of certain items, we have also intuitions about our practices of moral appraisal [Smith, M. 1994. p. 39-40 and Horgan, T. Timmons, M. 2006. p. 222-223]. As I will defend later, we sometimes refer to our own values and our moral commitments by presupposing their *truth*, by assuming that we *know* whether a certain action is right, or even by appealing to the *fact* that certain act is *clearly* wrong. We tend to acknowledge the fact that there is a certain link between our evaluative judgments and our will. And we are also well aware of our negative reactions when this motivational link is broken. Equally, when located at the center of an evaluative disagreement, we tend to use evidence to discard other evaluative opinions, securing our own moral commitments againts them. We do not simply pass the buck in these contexts, supposing that any opinion can be equally valid. Thus, it is an assumed fact about our moral practice that we endorse a set of

---

are, on the other side" Pettit, P. 2004. p. 306-307. Another attempt to characterize philosophy appears at page 308.

platitudes about our very evaluative activities, i.e. about the very *stance* that we occupy when we evaluate certain items as right or wrong, as desirable or undesirable, or as blameworthy or praiseworthy. We have intuitions about our own evaluative activities. They focus on the set of dispositions, emotional reactions, and patterns of social enforcement that underlie our evaluative assesments .

Having noted the centrality of these intuitions or platitudes about our moral stance, five important things deserve to be noted about them.

The *first point* to stress about our evaluative platitudes is connected with their theoretical *neutrality*. As I suggested above, our intuitions about our moral stance supply an observational basis from which *any* meta-ethical theory could be evaluated [Railton, P. 2006b. p. 205. Fischer, J. M. Ravizza, M. 1998. p. 10-11]. An important caveat is in place here nevertheless. To some people, the very formulation of our intuitions about our moral practice is implicitly supporting a definite image about the status of such practice – a realistic image, they tend to say. If this criticism has a grain of truth, however, it would jeopardize the role played by these platitudes in evaluating the accuracy of the differents meta-ethical theories. To put it simple, if our basic platitudes were presupposing a certain image about the status of our moral opinions they would not confom a neutral basis to determine the relative merits of the different meta-ethical theories. On the face of it, two options seem natural to me. Either we simply deny the methodological importance of our platitudes (because we assume that they are not neutral after all) or we are disposed to secure their methodological role by further securing their neutrality. I think that as far as our platitudes about our moral stance are very important to assess the merit of our meta-ethical theories, we should assume a general piece of advice about the proper specification of these platitudes. If I am right, it should be desirable to depict them along a minimalist line of commitment, i.e. without implictly presupposing any second-order view about the status of our moral opinions. We should be able to describe our intuitions neutrally, not assuming a priori any metaphysical view about the status of our moral practice. If we were able to do such thing, the regulative role played by our platitudes could be secured[2] [Blackburn, S. 1998. p. 80-83. Shafer-Landau, R. 2004. p. 24].

---

[2] Neil Sinclair writes: "(...) to say that moral talk is descriptive could be to say that moral talk involves the expression of mental states that descriptively represent the world in moral ways. And to say that moral talk is attributive could be to say that moral talk involves attributing moral properties (...) to actions, states of affairs and characters. Both these claims are part of the realist meta-theory. However, when we read these terms in this way then it becomes controversial whether description and attribution really are some of the pragmatically important features of moral practice. It is surely undeniable that moral talk involves moral predication (...) But to

The *second* point is concerned with the relative *importance* or *centrality* of our platitudes. As it will be clearer later, when I appeal to them to decide between the menu of meta-ethical theories, I am not assuming that every intuitive feature about our moral stance is equally important to discard one theory or another. Surely, some platitudes are far more important than others to encapsulate the inferential and behaviourial dispositions we tend to exemplify when we evaluate an item in moral terms. But if this is so, these constitutive platitudes are far more important in order to decide between opposite meta-ethical theories. Accordingly, our main meta-ethical theories should be able to accommodate, first and foremost, the *core* platitudes that define the image we have about our moral stance [Cuneo, T. 2007. p. 34].

The *third* point focuses on the type of access we have to our platitudes. Which level of knowledge do we have to pressupose in order to give sense to the methodological role played by these platitudes in shaping the meta-ethical debate? To answer this question I will stress a simple point. Let me start by assuming that we normally acccept that only a small number of speakers are able to articulate the inferential and behaviourial regularities that are fixing the content of a given concept (C). The possesor of C normally takes for granted several rules about C when he masters C. He takes for granted such rules because he acts *as if* he accepted the truth of those rules. These rules are somehow unfolding the behaviourial and inferential regularities to be exemplified by anyone who is using C correctly. In a certain sense, we tend to accept that the speaker does not need to be able to formulate these rules to use C across different contexts. But once this familiar image is reminded, what should we say about the access that a theorist who is trying to explain the linguistic behaviour of C should presuppose? Are we required to assume that the theorist has the ability to articulate the platitudes that shape his practice? Focusing on the moral case, the answer is clear to me. As far as our platitudes about our moral stance are usually referred to in a context of meta-ethical theorizing, we should demand from the meta-ethicist a certain ability to articulate the core intuitions or platitudes about our moral stance. It is because he can *explictly* articulate such platitudes that he will be able to articulate a maximally plausible meta-ethical view. And moreover, it is because we can offer such detailed inventory of our intuitions that we can use

---

assume that it also involves commitment to a certain meta-theoretical understanding of that predication is much more controversial. Better to say that the uncontroversial phenomenon here is the meta-theoretical neutral one of moral predication, with it then being a further issue how that predication is to be understood". Sinclair, N. (forthcoming)

them to decide between opposite meta-ethical theories [Railton, P. 1996. p. 66 for this general point].

Having stressed again the meta-methodological role played by our platitudes about our moral stance in shaping the meta-ethical debate, the *fourth* point I want to make moves around the way in which these intuitions help to resolve cases of meta-ethical disagreement. As we will see later, apparent disagreements between meta-ethical theories are sometimes simply caused by previous disagreements about the shape and content of our moral practice, i.e. about the set of intuitions that determine the nature of our moral stance or about the body of data that has to be explained by a second-order theory about our moral judgments. Thus, two meta-ethical theories (A and B) could favour two opposite accounts of our moral practice because they depart from two different set of platitudes. If this were so, a plausible way to resolve the disagreement between them should consider the importance of the disputed set of platitudes for an accurate understanding of our moral experience. At the very moment we establish that a certain platitude (P) is essential for an accurate understanding of our moral experience and that B does not accomodate P - while A does - we could claim that B is a less accurate than A in reconstructing our moral experience. Accordingly, the account offered by A should be preferred to the second-order reconstruction favoured by B [Gibbard, A. 2003. p. 18-20].

Finally, even if most of the times we tend to assume that our intuitive standpoint about our moral practices is directly accessible without adding any further revision or epistemic filter, the fact is that we should try to secure the epistemic authority of our basic platitudes or intuitions (at least if we want to use them to evaluate and sustain our theories about the status of our moral stance). But what sort of story could we offer to secure our appeal to intuitions in determining the shape of our moral practice?

There are several available options to offer a response to this general question. We could pretend, for example, that some individuals could get access to a realm of intuitive truths about the shape of our moral judgments. By following them, it could be argued, we could learn what are the basic features associated with a genuine moral opinion [McDowell, J. 1978]. Or we could assume, alternatively, that a bunch of sociological, psychological, or anthropological data could serve as basis to supplement our intuitions about our moral perspective. In a certain sense, simply by collecting and comparing empirically informed data about how different people, located at different contexts, are morally evaluating certain items,

we would be in a better position to grasp the basic features defining our moral stance [Nichols, S. 2004]. Finally, we could presuppose that any group of persons could enumerate a relevant set of intuitive features associated with the genuine utterance of a moral opinion once a certain process of epistemic and imaginative improvement - one focused on certain examples and thought experiments - has taken place [Pettit, P. Smith, M.1996. Jackson, F. 1998. Chapter 1].

Without entering into further detail, I believe that a sufficient route to secure our intuitions about the shape of our moral stance should focus on the last alternative. This route should reject any unqualified reference to a sort of peculiar capacity involved in grasping the core intuitions that structure the shape of our moral stance. Additionally, this route should escape from any unqualified sort of appeal to the intuitions of the majority in determining the relevant set of platitudes that fix the contours of our moral stance. Instead, the favoured route will assume that once a relevant set of cognitive and behavioural dispositions are isolated through a process of common deliberation we would be entitled to extract some central assumptions about the shape and content of such stance[3]. Any description of our moral stance respecting these general lines would be a plausible choice to sum up our intuitions about our moral stance.

In what follows I will pay attention to a set of features about our moral stance, i.e. about the stance occupy when we utter a moral judgment. I will respect an orthodox approach that distinguishes among the different types of platitudes in virtue of their formality or, equivalently, in terms of the degree of reflexivity required to connect such intuitions with our normal or unreflective moral practices.

---

[3] Importantly, I will not understand the set of features associated to moral judgments as classical conceptual items - as features of a more accessible nature to be added in the overall project to offer a conceptual analysis of the concept *morality*. In a certain sense, to identify these features is not simply to affirm that they are in the mind of typical users of the concept of morality. In a clear way, some of the basic features that fix the content of our moral practice are made explicit when we consider other general practices related with morality, such as politics and law. Besides, the complexity of some of the dispositions associated to these features makes it impossible to understand them as more simple conceptual items that are effectively accessible from merely a-priori or conceptual elucidation of the moral agent.

4. Basic platitudes about our moral stance

4.1 Formal Platitudes

Some of the most important features we are committed to when we occupy a genuine moral stance toward a certain item are *formal* in nature. It is not easy to precisely state what sort of claim are we endorsing by saying that some of our most pervasive intuitions about our moral perspective are *formal* but, as I understand it, when somebody assures that our evaluative perspective is pressing us to endorse a set of *formal* intuitions, what he is trying to identify will surely fall into four basic categories. Either (i) he is assuming a set of *grammatical* considerations about our moral practices, which are constituting any paradigmatic case of moral evaluation, or (ii) he is stressing the importance of a set of considerations related to the way in which such paradigmatic cases of moral evaluation can attain their precise moral status, or (iii) he is acknowledging a set of intuitions that highlight the *action-guiding character* connected with core cases of moral evaluation, or, finally, (iv) he is noting the social nature of our moral stance [Brink, D. O. 1989. p. 24-30. Smith, M.1994. p. 39-41. Timmons, M.1999. Chapter 4. Railton, P. 2006b. p. 202-205].

4.1.1. Platitudes about the surface grammar of our moral language

When we speak about our moral judgments, it has been customary accepted the pervasiveness of a grammatical intuition. It has been widely noted that when we adopt an evaluative or moral stance toward some item - action, institution, outcome or disposition - we *say* something about such item, i.e. we *assert* something about it [Blackburn, S. 1998. p. 70. Timmons, M. 1999. p. 128. Horgan, T. Timmons, M. 2000/2007. p. 58. Pettit, P. 2001. p. 236. Shafer-Landau, R. 2004. p. 24. Cuneo, T. 2007. p. 21. Sinclair, N. 2007. p. 344-345]. Keeping this intuition in mind, some philosophers have thought that two additional theses should be accepted in order to conceptualize, from a second-level stance, the nature of our moral utterances. On the one hand, we should accept that as long as our moral utterances say something about certain items (as they express an *assertoric* act of speech) they could be evaluated in terms of *truth* or *falsity*. On the other hand, we should convene that this assertoric function implies that our moral utterances are giving voice to a cognitive state, one whose

content is fixed by the just mentioned truth-conditions [McNaughton, D. 1988. p. 39 and Brink, D. O. 1989. p. 25. Jackson, F. Pettit, P. 1998. Timmons, M.1999. p. 144. Svavarsdóttir, S. 2006. p. 163].

   Although both second-order theses about our moral opinions are sometimes assumed as one along with the platitude about the assertive surface of our moral language [See Wright, C. 1992 for the platitude stressing the connection between *assertion*, *belief*, and *truth*], I would like to ignore them at this introductory stage. Or better; I would like to leave them aside if they are to be understood as theses derived from the surface grammar of our moral language. Even if we all assume that some moral opinions are *true* while others are clearly false, that we sometimes *believe* that some acts are wrong and others right, that we *know* that some moral commitments are better grounded than others, and that such knowledge can be sometimes *objective*, I will propose to accept instead a less ambitious interpretation of our grammatical intuitions about our moral opinions. In terms of my favoured choice, by noting that when we evaluate something in moral terms we are *saying something* about a certain item we simply assume that our moral language is mirroring *a family* of assertive or predicative uses. But by saying that we are *not* endorsing, in principle, any kind of commitment with a second-level explanation of this grammatical phenomenon. Although the surface grammar associated with our evaluative terms is taken to be similar to strong predicative or assertive uses, it implies nothing by itself about the semantic or psychological status of our moral opinions [4].

   Evidently, once this minimal interpretation is stressed, the grammatical appearances are not offering by themselves a conclusive reason in support of any given meta-ethical view but rather a topic-neutral fact, which has to be accounted by any meta-ethical account – even by

---

[4] The neutrality I am presupposing above is equivalent to deny that by focusing on the grammatical features of the sentence contained in the following utterance

   (1)   Torturing babies is wrong

   we can determine the type of speech-act exemplified by (1). By working out on such grammatical features we should not presuppose that the speech-act exemplified by (1) is assertoric, that is, one structured around the predication of a property (*wrongness*) on a certain act-type. Additionally, we should not assume, by simply focusing on the surface grammar of the sentence contained in (1), that by uttering (1) we express a belief whose content is *that torturing babies is wrong*. Finally, we should not presuppose that (1) is true in virtue of how the proposition expressed by (1) – *that torturing babies is wrong* – matches a certain fact, i.e. *the fact that torturing babies is wrong* – and that such fact is independent of speaker's attitudes. As far as these derivations conform the core of moral realism I am simply claiming here that we should not argue for such meta-ethical position by merely departing from certain grammatical features [Rosen, G. 1998. p. 395]. Thus, although the surface grammar of our moral language is assumed by all the meta-ethical parties, we should not derive from this basic agreement the chain of theses noted above about the status of our normative thought. (See Kalderon, M. 2005. p. 96-100, Cuneo, T. 2007. Chapter 1, and Wedgwood, R. 2007. p. 17-18 for three statements of this basic proviso. Compare again with note 1 again).

those meta-ethical views that defend a non-descriptive account of our moral discourse (see Chapter 2, section 2.3.1). The declarative or assertive function of our moral language should be considered a common intuition about our moral perspective, which has to be accepted even by those ultimately committed with an anti-descriptive or anti-realist meta-ethical framework [Smith, M. 1994. p. 39. Railton, P. 2006b. p. 207. Horgan, T. Timmons, M.2006. p. 222][5]

4.1.2 Platitudes about the kind of features constituting our moral stance

To characterize formally our moral stance it is not going to be sufficient to claim that it is centrally associated with a set of assertoric uses. It is so even if we further adduce our responsiveness to evidence in order to secure the assertive function associated with our moral opinions. Besides, we need to ask a simple question: in virtue of what additional features could those supposed assertions that compound our moral stance be separated from the wider set of non-moral assertions, i.e. from those that although responsive to factual information or evidence, do not determine or guide our action in any normative sense? Or to put in another way, what features are required to separate moral from non-moral assertions? I will argue that at least two basic features are sometimes appealed to in order to establish the basis for a distinctive type of moral assertion.

  Firstly, any appeal to evidence in the moral domain entails a formal constraint shared by the descriptive and the evaluative domains. In terms of this formal constraint, any assertion, if genuine, should be *consistently* guided by evidence. This formal constraint imposes by itself a general criterion to separate assertive from non-assertive uses and when it is particularized to the moral case it implies a general criterion that determines the shape of any moral assertion. In terms of it, if we are disposed to assert that a certain situation (X) is wrong, and this moral assertion is based on the behaviour of certain non-moral facts (N), we should be prepared to

---

[5] There are at least three additional facts in support of a wholeheartedly assertive-based interpretation of the surface grammar of our moral language. *First*, because of the fact that moral opinions are sometimes denied and also located into complex patterns of argument, it is sometimes assumed that those sentences uttered from our moral stance are exemplifying an indicative use, an use associated with the expression of a proposition [Blackburn, S. 1998. p. 71. Sinclair, N. 2007. p. 344]. *Secondly*, there are intuitions that suggest that our moral opinions exemplify an assertoric type of specch-act because we usually attach a high degree of importance to *factual information* to determine the quality of other's moral opinions [Boyd, R. 1988. p. 213. Brink, D. 1989. p. 202-203]. *Finally*, our assumption in favour of an assertoric or predicative model to understand evaluative utterances is surely driven by the way we resolve cases of *moral disagreement*. .

assert of every similar situation (X') that instantiates a similar set of non-moral features (N') that it is wrong.

This intuition is clearly supported by reflecting on our practice. If, for example, you affirm that a certain action (A) is wrong in a certain context (C), an ulterior question about the evidence supporting your claim in C is not entirely out of point: 'Why A is wrong in C?' Once you offer a set of non-evaluative considerations to support your assertion – once you offer certain evidence that has enough ground to assert that A is wrong in *this* case - some people assume that you are implicitly committed to a conditional claim with the following form: if you are confronted with a situation similar to C in all their non-moral features – call it C'- you should evaluate C' as wrong [Hare, R. M. 1952. 8.2, 10.3, and 11.5. Hare, R. M.1963 at 2.2, 3.1, and from 3.2 to 3.5. Hare, R. M.1981 6.1 and 6.4]. Hare located this formal feature under the general rubric of *universalizability*::

> "Moral judgments are, I claim, universalizable in only
> one sense, namely that they entail identical judgments
> about all cases identical in their universal properties"
> Hare, R. M. 1981. p. 108

It is important to note that *universalizability,* at least in the classical treatment of Hare, does not exemplify a special sense of *generality. Universalizability* refers, at least in Hare's sense, to a quasi-logical feature extracted from different situations where our acceptance of a moral claim could be fitted intuitively [Hare, R. M. 1963. p. 32, 35-37 Gewirth, A. 1960 for an opposite view about the meta-philosophical status of this feature]. As a result of the conceptual status of this intuition, in those cases where you are *not* disposed to assert that C' is wrong even when you were disposed previously to affirm that C was wrong – and C and C' are similar in respect with their general non-moral features – you should provide an explanation of the evaluative difference between C and C'. Maybe some features among those that grounded your judgment at t_1 are not operating in a right way at t_2; or maybe you have been reflecting further about the validity of the previous features, and have concluded that they are not valid to ground your moral opinions anymore. In either way, when C and C' are similar in non-moral terms but you are not evaluating C and C' in the same way, it is assumed that an explanation is not out of order.

Evidently not every likeness between two situations is relevant to determine the adequate basis for a genuine moral assertion [Railton, P. 2006. p. 205]. And this platitude is announcing the *second* thing that needs to be added to supplement the assertive picture referred above. Along with the purely formal platitude that stresses the importance of *consistency* in any assertive use, we usually acknowledge a further intuition about the *type* of considerations required to support a genuine moral assertion. Roughly, we assume that the non-evaluative features that we appeal to in order to make a moral opinion must be *impartial* and sufficiently *general* to be unaffected by contextual variations.

Impartiality is giving voice to a simple intuition. In terms of it, any non-evaluative feature that is adduced to justify a moral opinion should be independent of those interests associated with the particular individual, group or community that endorse the moral opinion. So, if I am endorsing a moral opinion I should be committed to appeal consistently to some non-evaluative features whose applicability is fully independent of my particular interest. In these cases in which I am departing from some partially-based attachments by offering a non-indexical feature as basis for my justification, the features being adduced are *impartial* or, so to speak, *generally relevant*[6]. Once we assume this further feature linked to our moral opinions, even when a moral assertion is governed consistently by the presence of certain non-moral features (as *universalizability* requires) we should assume that not every assertion supported by a consistent set of descriptive features is giving voice to a moral judgment. A further claim of *generality* or *impartiality* is needed to exemplify a genuine moral assertion.

4. 1. 3. Action-based platitudes

If we attended to the common phenomenon of moral motivation we could recognize a second subset of formal platitudes. Once we reflect on the inferential and dispositional commitment assumed for most in adopting a moral stance toward some items, it is assumed that there is a *reliable* connection between our moral opinions and commitments and those motives or states that guide our action. As James Dreier has recently noted, this connection between an agent's

---

[6] Note that I am accepting that a feature can be adduced as supporting *impartially* a moral judgment even if such feature is depicted in more or less agent-relative terms. If I assume, for instance, that some types of personal relationships are important *for me*, the only thing that *impartiality*, as a general constraint, requires is that the acceptance of such personal-based concerns can be defended as important for other people in whatever situations we can plausibly to consider. See Railton, P. 2006b. p. 202-203

moral assertions and his motives "is an alleged fact about ordinary moral thought, a datum for which every meta-ethical theory must account" [Dreier, J. 2006/2000. p. 547]. It would be a very strange fact, in sum, if you were judging that a certain action is wrong without showing any disposition to act in line with your moral opinion [Smith, M. 1994. p. 6]

To some people this *reliable* connection between moral assertion and action is so tight that sometimes it is assumed that even the very possibility of a genuine moral claim rests on the effective motivational import of the moral claim. They are happy to accept that the function played by our moral assertions will constitutively consist in guiding our own behaviour (or in correcting other's reactions and dispositions to act). Moral assertions, they argue, are conceptually connected with our motives. Geoffrey Sayre-McCord encapsulates this platitude along this compact statement:

> "When it comes to accounting for what thinking of something as good might consists in, apart from merely thinking it has some natural properties, people often appeal to the apparently intimate, and unique, connection between sincere moral judgment and motivation (…) To honestly think, of something, that it is good, they maintain, is ipso facto to have some motivation to promote, preserve, or pursue that thing. Conversely, to discover of someone that he is actually completely indifferent to what he claims is good is to discover that he does not really think it is good at all" Sayre-McCord, G. 2006. p. 49

But the thing is that, against the received opinion Sayre-McCord is giving voice, the accounts that equate motivational reliability with a kind of conceptual necessity only *interpret* our common intuition about the motivational import associated to our moral opinions. As I will defend later, the intuition that stresses the dynamic character of our moral commitments could be respected even if we were not defending a conceptual or logical connection between moral assertions and motives (chapter 3). The dynamic platitude, or so I will assume here, is basic in respect to any philosophical formulation of it [Shafer-Landau, R. 2004. p. 156-157]

Evidently, the just referred platitude is directly connected with a further intuition about the *explanatory power* associated with moral assertions. In terms of this connection, moral opinions could serve to explain human action because of the pervasiveness of the previous link between your moral assertion at t_1 and your motives to act at t_2. So, if you say at t_1 that 'X is the wrong thing to do' and I whish to explain your behaviour at t_2 - by pointing out a set of motives that will determine your behaviour at t_2 - a rational procedure would simply consider your moral judgment at t_1 to predict your action at t_2 [McNaughton, D. 1988. p. 20-24 and 106-116. Smith, M.1994. p. 7. Railton, P. 2004/1986. 2006b. especially at page 206. Sayre-McCord, G. 2006 p. 51-52]

The two intuitions that I have just sketched – *dynamic character* or motivational *reliability* and *explanatory relevance* – could be further connected with another central feature of our moral assertions. In terms of this further intuition, the motivational reliability linked to our moral assertions is closely associated with a set of emotional dispositions, patterns of emotional reactions, and feelings that are involved in our responsiveness to those (non-evaluative) features that govern the utterance of such assertions [Strawson, P. 2003/1962. p. 75].

That emotional responses and feelings are important in order to sustain core cases of moral evaluations has been stressed forcefully in recent times either by giving more credit to certain philosophical theories [Blackburn, S. 1998. Chapter 1. D'Arms, J. Jacobson, D. 2006 and 2000. Gibbard, A. 2006] or by being more responsive to certain psychological findings related with the emotional basis of our moral judgments [Damasio, A. 1994]. Let me assume those findings and focus again on a more intuitive level that comes from the inferential and judgmental dispositions involved in our normal practices. Once this perspective is assumed, some features point toward a sentimentalist account of our moral stance. Here I am going to stress three of these sentimentalistic-rooted features.

Let us consider, for instance, those cases where we endorse an apparently sincere moral judgment about a certain act but we are not further motivated to act in light of our moral assertion. These cases are fitted to tune our negative reactions toward the agent's moral failure, so we consider (besides of agent's proposed excuses) whether the agent is in the grip of some sentiments such as *guilt*, *remorse* or *regret*. Sentiments are relevant here, if not to determine the agent's responsibility, at least to modulate our reaction to his behaviour [Nichols, S. 2004. Chapter 1 and 5]

Additionally, it would be plausible to accentuate the importance of sentiments and feelings by stressing the obvious connection between the dynamic character of our moral judgments and the presence of certain sentiments and feelings. It could be pointed, for instance, that feelings and emotional reactions would enable the practical import linked to our moral assertions. In a certain way our emotional architecture should play a central role at some stage of the process by means of which our assertions are connected to our motives [Wallace, R. J. 2005. p. 107. Nichols, S. 2004 p. 18. Stich, S. Doris, J. 2005. p. 124-127]

But sentiments are paramount components of our moral life not only because they help to improve our moral evaluations by including themselves as relevant factor to be accounted for. Additionally, sentiments and feelings are relevant because of the role they play in order to structure and regulate our most basic processes of socialization and coordination. In these interpersonal contexts we tend to appeal to certain feelings and emotions, using them as primary tools in order to shape our responses to certain interpersonal-rooted problems (coordination, confidence, trust, etc). The idea to keep in mind for the present purposes is that, as far as the contexts where these problems of coordination emerge are quite pervasive in our social life, we are disposed to accept, without requiring a great amount of argument, the importance of certain emotional routes by means of which those responses that will offer a better pay-off in certain contexts are promoted or enforced.

4. 1. 4. Platitudes about the paramount nature and the relational character of our moral stance

The last family of formal platitudes that I want to highlight are those that refer to the *paramount role* or *importance* attached to our moral opinions when we deliberate about what to do and those that emphasize the *relational* dimension of those basic concepts included in our moral perspective.

Let me start with the intuition that stresses the *importance* or the *paramount character* of our moral opinions. Although moral judgments are salient in a phenomenological sense, the sense of *importance* I want to privilege here goes beyond this experiential level. In an obvious sense, to be under the grip of a moral obligation is not only to have a certain type of experience (a phenomenological urge to act in line with the content of our belief). In addition, to be under a moral obligation is to exemplify *a bunch of deliberative disposition*. We act in

light of a moral obligation when we are disposed to gather evidence in support of the required choice, when we devote a significant amount of time to deliberate about our moral requirements (integrating them into a wider space of obligations and duties), and when we are highly responsive to any factual mistake related with the applicability of the moral requirement. It is surely by considering the intensity or stability of such bunch of deliberative dispositions that we understand the importance of moral demands (and their deliberative priority in respect with another kind of demands - legal requirements, rules of etiquette, etc…).

What about the platitude that focuses on the *relational nature* of our moral stance? In the sense being assumed here, it simply points out the fact that those evaluations endorsed by an agent when he evaluates an item from a genuine moral standpoint are expressed by means of concepts coming from, at least, two basic realms. On the one hand, from a *deontic* realm[7] - *right* and *wrong* are the fundamental concepts here. From the *deontic* realm we can derive a secondary category of normative concepts such as *permissible*, *required*, *prohibited* or *supererogatory.* Thus, if an option is the only *right* alternative open to an agent in a certain situation, we usually claim that he is *required or obligated* to act in such way. An option is *prohibited* if it is required not to choose it. If an option is *permissible* then it is not morally wrong to perform it. If it could be *right* to opt for a certain choice but we are not required to choose it, we say that the option expresses a *supererogatory* demand. And so on [Smith, M. 2005].

But our basic *deontic* moral concepts presuppose certain mastery in the use of a related set of concepts aimed to discern those conditions where an agent can be *blamed* or *praised* for an action - where the action has been subject to evaluation under *deontic* categories. Thus, in cases where we evaluate whether an agent is *praiseworthy* or *blameworthy* by a certain wrong action (X) he did (because, let us say, a certain degree of *responsibility* or *agency* could be ascribed to his performance), some features associated with the capacities the agent is expressing in X-ing are forcefully appealed to in order to establish the moral merit of his action (by appealing to categories such as *control*, *skill*, *knowledge* of relevant *side-effects*, etc…). The sound link between our concept of *responsibility* and the general conceptual space

---

[7]I will distinguish here, following a standard use, between *deontic* or *thin* concepts, focused on *actions* or *choices* - 'right' or 'wrong' - and *evaluative* concepts, focused on dispositions, states of affairs, experiences and other sort of items -'good', 'bad' and their cousins 'desirable' and 'undesirable'. See recently Dancy, J. 2000. Chapter 2 and Smith, M. 2005

around the broad category of *intentionality* will require to connect our mastery with core moral categories at the level of *responsibility* with those conceptual abilities we express in those cases where we are applying a subset of intentional terms (*action*, *agency*, *intention*, *intentionality*) [Strawson, P. 2003/1962. Bratman, M.1987. Chapter 8. Fischer, J. M. Ravizza, M.1998. Chapter 3. Bennett, J. 1995. Scanlon, T. 1998. Chapter 6].

Besides the set of concepts above, in evaluating something in moral terms we normally appeal to another family of concepts. I am referring, of course, to *evaluative* concepts. They are sometimes alluded by philosophers either as referring to *thin* evaluative concepts such as *good* and *bad,* or as referring to *thick* evaluative concepts such as *pleasant*, *kind*, *brave*, and so on. Evaluative concepts, at least when appearing in moral contexts, are primary used to signal those items that make our life better. Character traits, institutions, experiences, and states of affairs are sometimes described as morally *good* or *bad*. And we describe some people as *kind* or *generous*. Evidently, what we have to ask regarding these evaluative uses is pretty clear: Could our common evaluative practices tell us something about the property – if any – that underlies all these uses? Unfortunately there is no clear prospect to reach an agreement on this point (at least not by merely considering the regularities underlying our evaluative practices). No clear platitude or set of platitudes about the content of our evaluative terms can be easily formulated at this level.

To many philosophers, however, there is something about our *evaluative* concepts that is pretty clear - even assuming the previous point about the limited power of our intuitive perspective to settle axiological issues. Our evaluative concepts are playing an essential role in determining the *rightness* of our actions. For some of them it is a quasi-conceptual truth, they tend to assume, that an act (F) is *right* if and only if F could maximize a certain *value*. Moreover, they assume that even if you disagreed on this particular intuition you should accept that the *rightness* of F is dependent on certain valuable features associated, either directly or indirectly, with F. Deontic concepts, they defend, can be understood in terms of evaluative concepts [Crisp, R. 2006. Chapter 1 offers a reduction going from the deontic (right) to the evaluative (well-being). See Smith, M. 2005. p. 13-19 for a less optimistic account].

4. 2 Substantive platitudes about our moral stance

Beside the previous formal features, a set of *substantive concerns* defines our moral stance. I assumed before that moral judgments should be understood, at least until further notice, as a sub-set of normative judgments. In this sense, *Epistemology*, *Legal Theory*, *Semantics*, and *Aesthetic* conform an unified realm of theoretical constructions, which share a common concern on the guidance of our actions (including linguistic action), our sentiments, and our most basic attitudes (*belief*, *desire*, and *intention*). Morality, in principle, should be located into this common normative domain. In all these domains a basic thought is shared: the guidance or authority offered by these norms is somehow related to reasons. Because our epistemic, moral, and prudential requirements are concerned with reasons, they guide our behaviour and mental functioning with the level of strigency that we usually presuppose to them [Gibbard, A. 1990. p. 40-41. Scanlon. T. 1998. Chapter 1]

Note that although this commonality of interests around the concept of *reason* could suggest, in principle, the existence of a shared terrain of agreement able to sustain a common characterization of our normative practices, the reality is that we usually define each of the domains included in the wider set of reason-based practices by isolating a substantive concern associated with each one of them. In line with it, sometimes it is defended that our beliefs (and the epistemic norms that are used to evaluate our cognitive performances) are constitutively linked to a certain epistemic goal - *truth* [Williams, B. 1971. Railton, P. 2004/1997. Velleman, J. D.2000a. 2000b, and 2000/1996]. A similar claim is sometimes made about our intentions [Velleman, J. D. 2000/1992. Raz, J. 1999. Setiya, K. 2006. Although see Frankfurt, H. 2004 for a different view]: they are all constitutively tuned in light of the satisfaction of a substantive concern - *good*. The interesting point here, however, is that if we assume a certain commonality around the concept of *reason* between our different normative domains, and additionally a certain degree of divergence in virtue of the different aims shared in each domain, a question comes immediately to mind: Which, if any, is the proper concern of morality once we have assumed that its authority is based on a certain type of reasons? What kind of substantive concern grounds such reason-based authority? And more importantly: How can we infer any substantive concern for morality from the incredibly complex set of inferential dispositions involved in our moral practice?

The answer to this is very clear. At a first sight, it could be suggested that while epistemology is basically concerned with a theoretical interest - 'what should I believe?' or 'what sort of principles can be formulated to maximize true beliefs?' or, alternatively, 'what sort of state are we instantiating when we know something?' - the standpoint we occupy when we reflect about our moral stance is basically interested in sorting out the reactions and responses of agents located in a *social context*. Peter Railton once wrote:

> "Moral evaluations seem to be concerned most centrally with the assessment of conduct or character where the interest of more that one individual are at stake (…) More generally, moral resolutions are thought to be determined by criteria of choice that are nonindexical and in some sense comprehensive" Railton, P. 1986. p. 21

At least from a philosophical standpoint, our moral judgments could be conceptualized as a type of assertions *aimed* to sustain our social or cooperative dimension. They do so by securing, controlling, and enforcing the reliable presence of some emotional and practical concerns on the agent. Taken as a normative practice, the institution of morality is aimed to coordinate and sustain our social dimension by inculcating certain beneficial sentiments enacted by some impartial considerations. The *reasons* appealed to by morality can be depicted as socially-rooted reason, reasons of "well-being, impartiality, and cooperation" [Railton, P. 1996. p. 63]. Thus, if someone claimed that a given moral norm is completely unrelated to the *well-being* of those affected by such norm - if he assumed that such norms are not related to certain impartial concerns - we should dismiss his opinion as completely implausible. By the same token, if someone proposed that the reasons offered by morality are esentially dependent of her own well-being – because they are constitutively related to her own partial concerns - we should assume that she is surely misunderstanding the sense in which moral norms give us reasons to act in a certain way [Joyce, R. 2006. p. 64-65].

Evidently, the precise elucidation of the way in which our moral practices could help to secure our shared social domain would depend on which theoretical concerns we share, i. e. on the type of conceptual commitments we embrace. For the present purposes, however, I would simply like to point that the conceptual elucidation of how the constitutive aim of morality (social coordination through impartiality) is achieved could be advanced by *Meta-*

ethics by offering an explanation of the kind of processes by means of which the basic goal is incardinated or included into the basic features defining our moral stance [see Joyce, R. 2006. Chapters 1 to 5]. Although the overall explanation that this requires would be complex, we can venture that it should move around (i) an identification of several temporally extended processes of enforcement - biologically rooted and aimed to secure the evolution of basic social institutions and arrangements - and (ii) a general hypothesis that connects, by means of a recourse to piecemeal explanations, such processes of enforcement with the central psychological features linked to our moral opinions [Blackburn, S. 1988/1993. p. 168-169. and 1998. Chapter. p. 8-21. Gibbard, A. 1990. p. 26. Nichols, S. 2004. Chapter 1. Sayre-McCord, G. 2006. p. 56-59. Brink, D. 2007. p. 271-272].

Let me finally refer to an intuition that, although aligned with our substantive commitments about the content of morality, could surely occupy a place among the formal platitudes related with our moral stance that were noted above. The reason why I will not locate this intuition at a formal dimension nevertheless is that I suspect that, while all the formal intuitions previously noted were widely accepted, the present platitude is severely contested from different fronts.

Before sketching it, let me return to the image that stressed the social character of our moral stance. Some philosophers are disposed to accept, in line with the social function of our moral assertions, that any decisional problem posited in moral terms should be resolved by appealing to a *consequentialist* procedure of decision. In every situation, they argue, the right option should be identified with the alternative, among those available to the agent at the moment of his decision, whose consequences are best in terms of a certain good (overall well-being, pleasure-satisfaction, preferences-satisfaction, ideal goods, etc…) [Scheffler, S. 1998/1985. p. 252. Nagel, T.1986. p. 164-180].

Philosophers attracted by *consequentialism* normally express their views along two possible dimensions not always coextensive. Either they assume a basic claim of *constitution* concerned with the sort of property that *rightness* denotes or they defend the validity of consequentialism as a *decisional procedure*. In the latter sense they affirm that a single rule or principle operates in every possible situation of choice: you should pick the best option, the one whose consequences are best) [see Hare, R. H. 1981. Kagan, S. 1994. Frey, R. 2002].

Evidently, some people reject the *constitutive* claim that establishes that *rightness*, understood as a property, is constituted by a promotion-based relation - one established

between an agent and certain states of affairs that are individuated in terms of *pleasure*, *well being*, or *preference-satisfaction*. They defend, in turn, that *rightness* is a sort of complex, second-level property, one constituted by a set of rules extracted from an agreement between those members of a society located in certain conditions of ideal information and mutual concern. When we claim that a certain action is right, they argue, we are simply referring to a second-order property, i.e. to a property attached to a certain act (X) in virtue of which X is permitted by any of the rules extracted from an agreement along the previously noted lines. As result of this opposition at the constitutive level, contractualists are disposed to accept a non-consequentialist decisional procedure, which assumes that in some situations the right action cannot be determined by weighing the goodness of the options (states of affairs) open to the agent but rather by asking for those rules that could be agreed upon certain conditions [Scanlon, T. 1998. p. 71-90, 116].

But the thing is that – and here is the intuition I want to stress at this point - although they are radically confronted, *consequentialism* and *contractualism* agree on a central image about the sort of access that we have to the relevant features that determine the rightness of an action. The image shared by both normative theories assumes that the access that we have to the features that constitute the rightness of a given choice must be necessary mediated by principles or rules - either by a single decisional *principle* or by a set of *moral rules*. However, an opposing intuition questions the centrality of moral principles in our moral reasoning. In terms of it, a moral decision does not require neither an unqualified sort of appeal to rules or principles to frame our moral decision nor *a posteriori* appeal to rules to justify of our choice [McNaughton, D. 1987. Dancy, J. 1993 and 2004. McDowell, J. 1998. McKever, S. Ridge, M.2006].

5. Looking forward

At the beginning of this chapter I assumed that any accurate characterization of *Meta-ethics* should divide its task into two separate stages. *Firstly*, it should take into account the nature of the *realm* or domain that is the focus of meta-ethics. *Secondly*, it should determine the type of *conceptual tools* required to offer a reliable explanation of the status of the opinions that are

embedded into such realm. Once we characterize *Meta-ethics* along this line, we must answer two basic types of questions before proceeding to extract meta-ethical conclusions:

(a) Questions about the nature of our moral practice, i.e. questions about the shape of the realm that is the focus of *Meta-ethics*.

(b) Questions about the conceptual components that are required to offer a second-order explanation of the nature and status of the moral realm.

In this chapter I have been mainly concerned with (a), and I have focused on some of the central intuitions that shape our understanding of the moral stance. I have suggested that once we accept the standard characterization of *Meta-ethics* as a second-order approach that is focused on our moral practices, pointing out the complexity associated to such practices will surely play an important meta-methodological role. In a certain sense, both the conceptual items compounding the second-level stance and the overall methodological ideal to be pursued by *Meta-ethics* (*accommodation*) should be influenced by the complex nature of the domain. Any meta-ethical approach, or so I want to argue, might take this complexity into account to modulate its methodological aims and its conceptual components in light of the complex nature of its object. I depicted the complex nature of our moral doman in the following way:
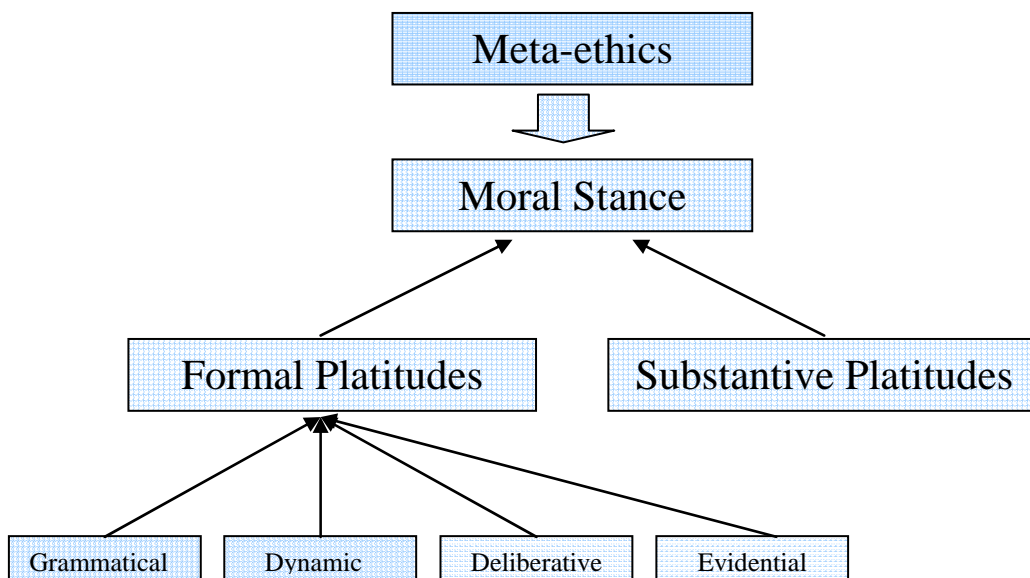
I assumed an important caveat about the diagram in section 3 above. There I noted that even if any meta-ethical theory should be concerned with the *accommodation* of the maximal number of intuitions about our moral stance, not every intuition should be considered as equally relevant to determine the explanatory merit of a given meta-ethical theory. Consequenly, I favoured a distinction between *constitutive* and *non-constitutive* platitudes. The methodological relevance of this distinction will emerge in the next chapter (Chapter 2, section 3). By now, however, let me simply sketch the three platitudes about our moral stance that I consider as somehow basic to determine our moral stance. If I am right, these platitudes have being the focus of the accomodation-project in recent *Meta-ethics*. The three platitudes are:

-   First, a set of features related to the *formal platitude* stressing the assertoric surface grammar of our moral language – *truth*, *knowledge*, and *objectivity*

-   Second, a *formal platitude* related to the *practicality* of our moral opinions

-   Finally, a *substantive platitude* related to the type of *normative guidance* being exemplified by our moral opinions – reason-based authority

If I am right, these platitudes are basic in defining the realm of facts that any meta-ethical theory should accomodate.

Leaving aside these issues around the centrality of certain platitudes, in the next chapter I will shift the focus of my interest. I will expose there how *Meta-ethics* – understood as a second-order discipline structured around a set of philosophical interests - has traditionally made justice to the complex nature of its focus. In a certain sense, I will be concerned with some methodological questions focused on the theoretical components required to offer an overall and plausible second-order approach on the first-order realm depicted above. In terms of the previous partition I will be concerned with some issues related with (b) above.

Chapter 2.

The Accommodation Project in Perspective

1. **Introduction.**

2. **The progressive widening of *Meta-ethics***

    2.1 Moore's classical view about the content, method and scope of *Meta-ethics*

    2.2 The intermediate stage of development of *Meta-ethics*

    2.3 Contemporary *Meta-ethics*

3. **The *impasse* in *Meta-ethics***

4. **Looking back and looking forward**

Chapter 2

1. Introduction.

In the previous chapter I suggested that *Meta-ethics* offers a second-order analysis of our *moral stance*, that is, a second-order account of the standpoint from which we use our basic moral terms. Straight as it is, I stressed in the previous chapter an important caveat about this general characterization of *Meta-ethics*. I noted that the first-order standpoint on which *Meta-ethics* focuses is not a uniform or simple domain. Rather, the first-order moral domain is structured around a set of platitudes that define its shape and content. Thus, I tentatively ended by assuming that *Meta-ethics* should offer a second-order analysis of our moral perspective *as it is being unfolded by a certain set of platitudes about our moral experience*:

Nonetheless, *Meta-ethics* can be further characterized by adding three additional levels of analysis. *First*, we should determine better the relative importance of the second-order disciplines that define the core of *Meta-ethics*. I will refer to this level as the *content* of *Meta-ethics*. Although we sometimes consider *Meta-ethics* as offering a *static* second-order account of our moral experience (a second-order account focused on *metaphysical*, *semantic*, *motivational*, and *epistemological* questions about our moral opinions), the truth is that *Meta-ethics* has not always been defined around the same set of second-order disciplines. The perspective from which *Meta-ethics* deals with questions about our moral stance has evolved historically. For this reason, a fundamental dimension of analysis to understand what *Meta-ethics* is about should focus on the synchronic or historical process by means of which its analytical core (*epistemology*, *ontology*, *semantics*, and *philosophical psychology*) has evolved.

S*econd,* we could improve the characterization of *Meta-ethics* by highlighting a set of methodological developments that aim to get some philosophical answers about the status of our moral opinions. I will refer to this realm as the *methodological* side of *Meta-ethics*. What sort of methodology is proper of *Meta-ethics*? Are current meta-ethicists still endorsing an *a priori* approach to second-order questions about morality or are they committed, on the contrary, with the prospects of an *empirical* explanation of our moral stance? In this chapter, I will assume that we should answer these questions to be able to characterize *Meta-ethics* beyond the general and schematic picture depicted above. In doing so I will offer an historical

overview of the methodological development of *Meta-ethics* and a brief sketch of the alternatives open to *current Meta-ethics*.

Finally, we could get a better grasp of what *Meta-ethics* is about by answering some questions about the relationship between *Meta-ethics* and *Normative ethics*. The *scope* of *Meta-ethics* will be presented when I discuss the type of relationship that can be established between meta-ethical insights and more substantive views about what is right or wrong in a moral sense. As I suggested in the previous chapter, sometimes *Meta-ethics* has been conceptualized with entire independence of normative theory. According to this common position, second-order claims do not imply any view about the deontic status of actions or institutions, i.e. they do not imply any first-order theory about the substantive focus of our moral thoughts [McNaughton, D. 1988. p. 163-175]. Although it was widely established for several decades, this common intuition has been recently subjected to several criticisms, most of which are mostly inspired by John Rawls's pioneering work on moral methodology [Rawls, J. 1951]. Surely, a good characterization of the content and shape of current *Meta-ethics* should take into account the historical development of these concerns. Or so I will argue.

In essence, we could achieve a more complex image of *Meta-ethics* once these additional dimensions are taken into account:
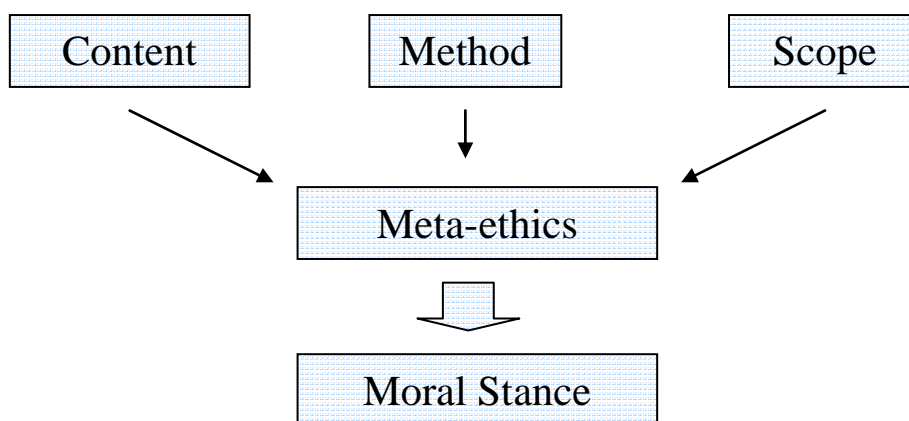
The current chapter has two main aims. First, I will offer a historical account of *Meta-ethics* by summarizing the evolution of this second-order discipline along the three dimensions noted above. After this, I will explain in what sense current *Meta-ethics* is facing a sort of *impasse*.

2. The progressive widening of *Meta-ethics*[8]

In this long section I want to illustrate how the classical view about the nature of *Meta-ethics* - which I will identify with Moore - has been replaced, during the last fifty years or so, by a more complex image of the aim and method proper to *Meta-ethics*. I will present three stages in the development of *Meta-ethics*. First, I will present Moore's own view of *Meta-ethics* as the initial image of *Meta-ethics*. In my opinion, Moore's influence spans from 1910 to 1930 (section 2.1). Second, I will explain how, from 1930 to 1970, Moore's classical image shifted progressively toward a more integrative account of the content, method, scope, and presuppositions of *Meta-ethics* (section 2.2). Finally, I will introduce some contemporary developments in *Meta-ethics* going from 1970 until now (section 2.3).

2.1 Moore's classical view of *Meta-ethics*

There is a wide agreement about the importance that Moore had in defining the shape of *Meta-ethics*[9]. Moore developed his particular conception about the limits of *Ethics* as philosophical discipline in several publications. Notably, the first chapter of *Principia Ethica (PE)* - 'The Subject Matter of Ethics' - has been traditionally pointed out as the *locus* of the beginning of *Meta-ethics*[10].

---

[8] The image that I am going to present rests on some influential accounts of the development of *Meta-ethics* during the last century. The structural image favoured here, for instance, owes a lot to a classic paper by Frankena (Frankena, W. 1951) and to a more recent piece by Stephen Darwall, Allan Gibbard, and Peter Railton (Darwall, S., Gibbard, A., Railton, P. 1992). In the same vein, the first chapter from Timmons's *Morality without Foundations* has helped to me contextualize the idea of *accommodation* into a wider historical perspective (Timmons, M. 1999). McNaughton offers a very good overview of the aims and problems faced by *Meta-ethics* in McNaughton, D. 1988. Equally valuable are the work of Railton, P. 2004/1986 and 1990, Wallace, R. J. 2006/1990 and Smith, M. 1994. Chapters 1-4. Shafer-Landau, R. 2004 and the introductory pieces by Shafer-Landau and Cuneo contained in Shafer-Landau, R. Cuneo, T. 2007 have been also helpful.

[9] For a recent example, see some of the papers devoted to Moore in the 2003 Symposium hosted by *Ethics*, the rich and multi-level assessment of Horgan, T. Timmons, M. 2006 and the pieces contained in the second half of Nuccetelli, S. Seay, G. 2007

[10] There are other references to find more insights on Moore's conception of philosophy as, for example, a piece of 1922 entitled 'The Nature of Moral Philosophy' (Moore, G. M. 2005) or the replies of Moore to his critics (Moore, G. E. 1942). As I will mainly focus on Moore's initial formulation of the content and scope of *Meta-ethics*, I will concentrate here on the material in Moore, G. E. 1993/1903 and 2005/1912. See Hurka, T. 2003 and Sturgeon, N. 2003 for two recent overviews on Moore's initial development

In the initial paragraphs of PE, Moore assumes that *Ethics* - what he took as 'an ideal of ethical science' - can be divided into two main parts. First, there are *semantic questions* about the meaning associated with our moral terms and with the moral sentences that contain them. Moore argued that these questions mainly focus on the possibility of finding definitions of these moral terms and of the sentences that contain them by means of some set of non-moral terms. That is, whether the *concepts* expressed by sentences containing moral terms -'what the object or idea (…) that the word is generally used to stand for' [(PE, 6) and Moore, G. E. 1922. p. 136] - could be expressed by using another expressions that not contain any moral terms as constituents. In Moore's terms, *Ethics* should deal with a basic question: Can moral basic concepts be expressed by means of definitions using non-moral terms? [Timmons, M. 1999. p 15-21. Horgan, T. Timmons, M. 2006. p. 1].

In addition, Moore defended that there is a second class of issues that should concern *Ethics*. These questions are not related with the *meaning* of moral terms, but rather with the particular items that exemplify the sort of properties denoted by moral terms. For example, the actions that are right, the things or characters that are good, or the laws and edicts that are just [Moore, G. E. 1993/1903. p. 54]. Moore was extremely clear about the kind of relationship to be established between semantic and more substantive issues about morality:

> "Ethics is undoubtedly concerned with the question what good conduct is; but being concerned with this, it obviously does not start at the beginning, unless it is prepared to tell us what is good as well as what is conduct" Moore, G. E. 1993/1903. p. 54. 6

Equally, Moore was clear about the type of approach needed to ask semantic questions about our moral claims. In his terms, the *semantic* level should be constructed around a *conceptually* oriented practice, which should determine the true of any proposed definition of a moral term without any appeal to empirical evidence [Soames, S. 2003. p. 39-41]. The scope that Moore presupposed for semantic questions inside *Ethics* can be encapsulated by the kind of definitions of moral terms that he proposed to analyze the semantic content of our moral opinions. Moore was interested in *reductive analytic definitions*. Given a moral term or expression M and a non-moral term or expression N, a sentence D_N expresses a *reductive analytical definition* of M in terms of N if and only if (a) D_N is a member of a set of S-

sentences establishing *synonymy relations* whose general form is 'M = D$_N$' and (b) the truth of the relation of synonymy can be established merely by considering the meaning of the terms and expressions contained in M and D$_N$[11] [Moore, G. E. 1993/1903. p. 57]. Thus, Moore proposed that *semantic* questions focus on the availability of a *reductive* restatement of the meaning of our moral terms. The merits of such reductive definitions could be assessed by means of conceptual analysis. That is, by reflecting on our intuition about the logical or grammatical behaviour of those non-moral concepts referred to in the reductive definitions[12].

But besides accepting the two previous claims (the one about the s*emantic-oriented content* of *Meta-ethics* and the claim about the *a priori methodology* to be used in the philosophical branch of *Ethics*) Moore was committed with two additional statements. In his terms, a question like 'What is X?' should be understood in a literal way, i.e. as asking something close to 'What sort of *property* X stands for?' Questions about the meaning of *good* should be understood as ontological-rooted questions, which try to isolate the sort of property that moral terms stand for. I will refer to this additional assumption as *the ontological-based view of conceptual analysis* [Dreier, J. 2006. p. 206-207]

Finally, Moore accepted a methodological view about *Meta-ethics*. He assumed that *semantic* concerns - including both *epistemological* concerns about the sort of justification associated with moral claims and a set of *ontological* thesis about the status of moral terms [Moore, G. E. 1993/1903. p. 34] – should be pursued *independently* of any question about what concrete items are good, or what actions are right[13]. This methodological insight, however, was further narrowed by Moore by presenting it as *the thesis of instrumental primacy* of *Meta-ethics* (the *semantic* branch of *Ethics* in Moore's terms) in respect to *Normative ethics* (the *substantive* branch of *Ethics* in Moore's terms). So, while *the claim of independence* cited above accepted the possibility of disentangling semantic and normative issues, *the thesis of instrumental primacy* assumed that semantic issues should be approached previously than any pursuit at a substantive level. Semantic issues, Moore argued, are

---

[11] By using Mark Timmons's words, reductive analytic definitions were aimed to offer 'an illuminating, compact and complete analysis of moral terms by means of non-moral sentences' See Timmons, M. 1999. p.18

[12] To reflect on whether such non-moral concepts could be substituted, without any loss of sense, in processes of moral reasoning or, alternatively, in situations of moral disagreement See Moore, G. E. 1906. p. 34 and the preface to the first edition. For this general point see Gibbard, A. 2003, especially Chapter 2. There, Gibbard presents a positive assessment of Moore's insights appealing to some cases of *conceptual substitutivity* around normative disagreements.

[13] Moore argued that neither in *selecting* the subject matter to be analyzed (moral language) nor in *evaluating* the success of the proposed semantic analysis, it would be required to appeal to our normative judgments. See, for a rejection of this thesis, Gewirth, A. 1960, especially at p. 169 and Darwall, S. 2006. pp.19-22 and 34-36. Smith, M. 1994, pp. 1-2 supports, among many others, a certain claim of independence.

philosophically more fundamental than other issues related to the assessments of several items in terms of goodness or rightness. In referring to the instrumental importance of *Meta-ethics,* he wrote:

> "It is an enquiry to which most special attention should be directed; since this question, how 'good' is to be defined, is the most fundamental question in all Ethics (….) Its definition is, therefore, the most essential part in the definition of Ethics; and moreover a mistake with regard to it entails a far larger number of erroneous ethical judgments than any other." Moore, G. E. 1993/1903. p. 57.
> **5**

In sum, Moore was committed with four basic claims about *Meta-ethics*:

(1) A claim about the ***content***: *Meta-ethics* is mainly concerned with *semantic* issues related with the meaning of moral terms plus some *epistemological* problems.

(2) A claim about the ***methodology***: *semantic* analysis is carried on a priori basis pursuing to establish a *reductive analytic definition* of moral terms.

(3) A claim about the ***ontological import associated with the semantic analysis***: 'What is X?' is analytically equivalent to 'What sort of property X stands for?'

(4) Two general corollaries associated with the scope of *Meta-ethics*:

- The ***independence*** of *Meta-ethics* from *Normative ethics*

- The ***priority*** or, better, the ***instrumental priority*** of *Meta-ethics* in respect of *Normative ethics*

In what follows, I will presuppose that (1), (2), and (4) conform the core of Moore's conception of *Meta-ethics* –(3) can be included into (1) if we keep in mind the semantic

theories that surrounded Moore around 1903. Thus, I will use these basic points to sketch the historical development of *Meta-ethics*. Let me start by introducing some initial reactions to Moore's conception of *Meta-ethics* coming from non-cognitive theories of morality.

2.2 The intermediate stage of development of *Meta-ethics*

I have divided Moore's meta-philosophical conception into three main claims: A claim about the *content* of meta-ethics, a claim about the methodology to be used to formulate and resolve meta-ethical disputes, and a claim about the *scope* of meta-ethics, i.e. about the kind of relation to be established between *Meta-ethics* and *Normative ethics*. In the next sub-sections I will adopt this division to offer some clues about the evolution of *Meta-ethics* between 1930 and 1960[14]

2.2.1 From a semantic-based view of the proper object of meta-ethics to a psychological-based account of the nature of moral judgments.

The progressive widening of *Meta-ethics* was firstly articulated around the rejection of Moore's semantic approach in favour of a more psychologically rooted model of theorizing. This general movement was favoured mainly by non-cognitivists theories. However, it is important to note that before such widening took place, the meta-ethical scene was characterized by the prevalence of some form of non-naturalism. It could be assumed then that after the addition of non-cognitivism to these established landscape, three possible meta-ethical choices conformed the range of options available to explain the nature of our moral opinions: *non-naturalism*, *naturalism*, and *non-cognitivism*. The orthodox tale assumed that this coexistence was broken around 1950 by a clear dominance of non-cognitivism over the other possible options. Whilst this dominance has been exaggerated, the fact is that non-cognitivism was imposed as the most solid position in *Meta-ethics*, or default view, or simply as the account that could accommodate better the argument established by Moore's open

---

[14] Three general points must be kept in mind to understand the sort of oscillations proper to meta-ethics after Moore's initial statements. *First*, the evolution along these three realms was, to a great extent, an *interdependent* sort of evolution. *Second*, the change along the three dimensions was very *gradual*. And *finally*, as a consequence of some features previously suggested, the process shaping the form of meta-ethics was *inclusive*. As consequence, any reconstruction of this period must end by offering an image more inclusive than those favoured by Moore.

question. It is against this general framework that we should understand the process of enrichment that I am going to expose in what follows. Let me start by Ayer's movement.

Ayer's approach

Ayer famously defended that our moral opinions must be understood as linguistic devices that express *emotional* states or *pro-attitudes*. In Ayer's view, when we endorse a moral opinion we are neither *describing* the reality in moral terms (as the non-naturalists defends) nor *representing* any amount of natural information by means of a moral predicate (as naturalism sometimes assumes). On the contrary, when we judge that something is right we are *expressing* an emotive state, a sort of *moral* feeling or *moral* emotion that is focused on the item that is been evaluated. As a result of that moral opinions, Ayer argued[15], cannot be assessed in terms of truth or falsity [Ayer, A. J. 2001/1936. p. 104, 109 and 171][16]

This reaction to Moore's non-naturalist cognitivism was accompanied by a methodological shift. In Ayer's framework, *Meta-ethics* should be concerned with a negative task. It should offer an explanation of why ethical concepts are, in fact, pseudo-concepts, that is, concepts whose function is entirely outside the cognitive realm. Ayer's statements in this sense are fairly famous:

> "We find that ethical philosophy consists simply in
> saying that ethical concepts are pseudo-concepts and
> therefore un-analysable (…) There cannot be such a thing
> as ethical science, if by ethical science one means the

---

[15]To some people, Ayer's own theory of truth (a redundancy theory of truth) could have made available a sense in which moral sentences can be assessed in terms of truth or falsity. See Dreier, J. 2004. p. 24

16 Ayer's emotivism faced two kinds of problems: *The first problem* was the inherent implausibility of the semantic constraints associated with *emotivism*. The principle of verification, as it was formulated by Ayer, was clearly implausible from the very beginning because it did not offer a criterion to separate meaningful statements from nonsensical ones. See Smith, M. 1994. p. 21-22. Miller, A. 2003. Ch. 2. Soames, S. 2003. *The second problem* was directly associated with Ayer's main idea, viz. that when we make a moral claim we are expressing a feeling or a sentiment. It is difficult to trace a sharp line between non-cognitive expressions of an emotion and the factual conditions required to express an emotional state in a minimal rational way. See Blackburn, S. 1984. p. 159-160. Jackson, F. Pettit, P. 1995. 1998. In relation to this point, emotivism faced another problem: Ayer's theory did not specify a clear criterion to separate non-moral expression of feelings from moral ones. So, in a certain sense, a mere expression of a sentiment could be in principle equated with the expression of a genuine moral opinion.

elaboration of a 'true' system of morals". Ayer, A. 2001/1936. p. 116

By establishing that our moral predicates are not representative of any portion of reality, Ayer motivated a new route that future meta-ethicists would follow. *Meta-ethics* should be necessary connected with a set of *psychological* questions. Ayer assumed that this was a natural consequence of assuming a non-representational story about our moral predicates [Ayer, A. 2001/1936. p. 116]. This methodological shift was very important for the future of *Meta-ethics.* After Ayer's initial suggestion, the core of the discipline became dominated not only by a *semantic*-based approach but also by a general interest in psychological questions related with the kind of thought that is expressed in uttering a moral opinion. The classical core of *Meta-ethics* (semantic, ontology, and epistemology) was supplemented, in sum, by the addition of a *psychological* interest focused on the nature of our moral thoughts [Gibbard, A. 2003. p. 6-7].

Despite of Ayer's radical views, the downfall of the classic program of ethical analysis was a very slow process. The full acceptance of its insufficiency was connected to several factors and their effects were not always direct for *Meta-ethics.* After Ayer's initial insights, the classical program of analysis - besides of being supplemented by a psychological concern - was widened by a more pragmatic approach based on the analysis of the contexts where we use moral terms. Thus, by assuming a theory of meaning focused on the *use* of moral terms in everyday contexts, pragmatic-oriented philosophers like Stevenson favoured an alternative approach *Meta-ethic*s. As a consequence, by the late forties *Meta-ethics* was mainly concerned with psychological and pragmatic issues about morality, abandoning a narrow semantic concern with moral language. Stevenson's meta-ethical position offers a good sample of this movement.

Stevenson's pragmatic turn

To study the behaviour of moral terms, Stevenson favoured situations of *moral disagreement* [Stevenson, C. L. 1944. p. 2][17]: Like Ayer, Stevenson analyzed these situations by taking a

---

[17] Stevenson's 'working model' to analyze any case of disagreement moves along the following lines:

A and B disagree about X if and only if:

straight psychological stance. In doing so, he accepted an heuristic principle: To elucidate the type of psychological attitude that is involved in the contexts in which two persons disagreed, it would be necessary to pay attention to the kind of evidence being used to resolve the disagreement. In Stevenson's sense, this type of evidence could help to determine the identity of the psychological state expressed by a moral opinion and the type of meaning associated with the terms included in moral sentences. Thus, the meaning and truth-conditions of our moral judgments could be established through a pragmatic process of inquiry about the dynamics of disagreement contexts [Stevenson, C. L. 1944. p. 20, 26, 82-84].

To put this pragmatic-based intuition to work, Stevenson established a conditional test. If the processes involved in resolving a disagreement between A and B are aimed at getting some additional factual evidence, then we should concluded that the states being expressed are beliefs. If, on the contrary, the processes involved in resolving a disagreement are not aimed at getting additional evidence, then we should assume that the states being expressed are not cognitive. By focusing on certain examples - *working models of analysis* – Stevenson established that the antecedent of the first conditional is false in those cases where the disagreement is about an evaluative issue [Stevenson, C. L. 1944. p. 11][18]. As a result of it, Stevenson suggested that non-cognitive processes were central to resolve cases of moral disagreement. Even if they are importantly connected to cognitive processes, i.e. processes based on the tracking of evidential information, Stevenson noted the centrality of non-rational processes of persuasion in cases of moral disagreement[19].

Besides presenting a substantive view about the meaning of our moral terms, Stevenson implicitly assumed, like Ayer did, a methodological intuition about the proper standpoint to do *Meta-ethics*. In a nutshell, he argued that by focusing on the *pragmatics* of disagreement we could get a fundamental level of insight about the *psychological* nature of our moral

---

(1) "X is wrong", as uttered by A, means *I disapprove this*; *do so as well*.
(2) "X is right", as uttered by B, means *I approve this*; *so do as well*.

[18] Note that Stevenson did not endorse an unqualified non-cognitive theory about the meaning of moral terms - one defending that our moral judgments are simply expressing emotions. He accepted, on the contrary, a more qualified version of non-cognitivism. In Stevenson's terms, even if we accepted that our moral assertions are primary directed by a commendatory function, the cogency or pervasiveness of our assertions is somehow determined by the evidence presented in the moral arguments adduced to resolve cases of disagreement

[19] In Stevenson's sense, the beliefs adduced to cancel a disagreement are denominated 'supporting reasons'. The role attached to *supporting reasons* is enabled by the causal relationship between facts about an agent's situation and facts about his desires and goals. When this relation is acknowledged, it is normal to expect that when we are in a situation of moral disagreement some beliefs could change our attitude. See Stevenson, C. L. 1944. p. 28, 115-118, especially at 114.

utterances in *any* context (see chapter 7 for a more detailed discussion of this general strategy). Although still entirely located on the methodological framework offered by the classic conception of analysis, *Meta-ethics* started to search for additional options besides of a narrowly construed analysis of moral terms in Stevenson's hands. Thus, by the time Stevenson's main work came out (1944), the content of *Meta-ethics* was definitely biased toward psychological issues - more precisely: toward psychological topics *pragmatically* framed. Following the path opened by Ayer and Stevenson, many people refined and completed the rejection of the classical view of *Meta-ethics*. Among them, no other was more prominent than Richard Hare.

Hare's program

Although Hare's overall image of the meaning and function of our moral predicates was congenial in its basic lines with the proposals defended by Ayer or Stevenson, it was backed by a more detailed picture of the dynamics underlying our moral practices. Hare noted that the analysis of our moral terms should meet two basic desiderata. First, any account of the set of conditions determining whether we are entitled to apply an evaluative term should be guided by a simple logical constraint: judge similar cases in a similar way - *universalizability*. In Hare's sense, *universalizability* constrains the rules that govern our use of non-evaluative terms. Second, any meta-ethical account of the meaning of our moral terms should explain how we should accommodate the *prescriptive character* of morality, the way in which evaluations are consistently related to the guidance of our behaviour[20].

Hare stressed two basic points. On the one hand, he assumed that both constraints - *universalizability* and *prescriptivity* - are *not* independent. We cannot fully master the conditions where an evaluative term can be *consistently* applied without grasping its prescriptive or action-guiding character [Hare, R. M. 1963. 2.5 and 2.6]. On the other hand, Hare defended that even if descriptive features are usually central to justify our moral opinions [Hare, R. M. 1981. Chapter 6], the *primary meaning* of our moral terms is determined by its prescriptive or action-guiding role in moral discourse, i.e. by the guidance

---

[20] For the first treatment of *prescriptivity* see Hare, R. M. 1952 section 11.2. For Hare's earlier account of *universalizability* see Hare, R. M. 1952. 8.2, 10.3 and 11.5. See Hare, R. M.1963. 2.2. 2.7. 3.4 and 3.5 for a further treatment of this logical feature and 2.2, 2.8 and 3.3 for a new look on the *prescriptive* character associated with our moral terms. The introduction to Hare, R. M. 1981 offers a general statement on the importance of logical analysis for moral philosophy

of one's own action *in light of universalizable features* from one's own deliberative perspective. In Hare's sense:

> **(Hare)** 'X is good', as uttered by me = I universally prescribe a favourable attitude toward X[21]

But leaving aside by now Hare's own view on the nature of moral judgments, did he endorse the psychological and pragmatically oriented shift championed by Ayer and Stevenson? To my knowledge, Hare's stand on this issue is one of scepticism. He thought that he central task of *Meta-ethics* could be fully achieved by means of logical or linguistic analysis. In Hare's sense, linguistic analysis is aimed at highlighting the set of criteria that would permit or secure the assertion of a normative claim. Thus, even when Hare recognized at several places that it would be desirable to ask for the psychological import of our moral opinions, the real fact is that he did not pay much attention to this level of inquiry to articulate his own approach[22].

Realism (Foot and Geach)

Although non-cognitivism marshalled *Meta-ethics* for a long time, during the mid-fifties some realistic alternatives were developed to resist the irrealist control. Among the most important additions, Philippa Foot and Peter Geach's neo-aristotelian proposals were fundamental to

---

[21] It is important not to read Hare's analysis as a straight definition of the meaning of our moral terms, in line with Moore's methodological program. Hare's approach was a partial departure of this methodological ideal. Timmons and Horgan quote to Hare as defending that "it is not part of my purpose to 'reduce' moral language to imperatives" [Hare, R. M. 1952. p. 2]. After that they write: (…) his meta-ethical position [Hare's] can be called weakly reductive in its use of non-assertoric, non-declarative sentences, as the model of moral thought and language, even when it eschews the strongly reductive claim that moral utterances are synonymous with, or semantically interchangeable with, imperatives". Horgan, T. Timmons, M. 2006b. p. 259. See for this general point Horwich, P. 1993.

[22] Two reasons can be adduced to support Hare's negative. The *first* reason is not directly obvious and it is related with assumption coming from the Philosophy of Mind. In Hare's time – his first book, *The Language of Morals*, was published in 1952 - there was a wide agreement around a behaviouristic sort of theory of the mind. In a nutshell, behaviourism had a negative stance toward any approach presenting an image of the mind where any kind of internal mental states were causally and explanatory relevant to analyze human behaviour See, for example, Hare, R. H. 1952. p. 165-166. The *second* reason against a psychological approach was even stronger. In the end, Hare was completely convinced that a program of logical analysis well carried was not only sufficient to offer an account of the meaning of moral terms but it should also be auto-sufficient and *more easily* stated in relation to other sorts of analysis. Hare wrote: "Since what we are discussing is the logic of moral language and not the tangled subject known as moral psychology, I shall not here inquire further into the fascinating problems discussed by Aristotle (…) Hare, R. M. 1952. p. 169

sustain the centrality and maybe the philosophical respectability of moral realism [Foot, P. 2002/1958. 2002/1958-1959. 2002/1961. Geach, P. 1974/1956].

For these realists, the core conceptual tenet of non-cognitivism moved around a simple intuition: although intimately related with the factual realm, the use of our moral terms (and so the argumentative patterns in which they are sometimes included) is governed by a set of rules or principles that cannot be understood as intrinsically tracking any amount of evidence. Evaluative utterances are grounded on a variety of reactions exemplified by the person uttering the evaluative judgment. These reactions, although modulated by descriptive-based rules, cannot be understood as intrinsically referring to any realm of descriptive facts [Foot, P. 2002/1958. p. 99. 2002/1958-59. p. 112]

Realists at the mid-fifties attacked this core intuition by focusing on simple cases of evaluation, i.e. cases where we say that a knife is good, or that certain person is a good farmer. They extracted some important consequences from the analysis of these simple cases [Foot, P. 2002/1961. p. 145]. Among the most significant, realists noted that when we say that a certain item is good there is an internal connection between a set of good-making features - a set of natural features of the object whose relevance or salience is fixed either by a functional criterion or by social standards of excellence - and the correctness of our evaluative use. In saying that a knife as good, for instance, the speaker is taking into account certain rules that determine the sort of evidence required to apply *good* when faced with a knife [Foot, P. 2002/1961. p. 137].

In some way, the semantic theory defended by moral realism in those years was very close to the following statement:

> **(Realism)** 'Y is good', as uttered by A, expresses a genuine evaluative use - 'good $_{prescriptive}$ of Y' - if and only if Y meets a set of N-features attached to Y-types and A believes that Y' falls under the set composed by Y-types [Foot, P. 1958. p. 113-114]

Moral realists assumed that the rules that govern the use of basic evaluative concepts are internally connected with certain *rules of evidence*. Rules of evidence, Foot argued, guide the use of an evaluative concept and the emotional response linked to such concept. Or to put the point in the terms favoured by Peter Geach: *Good* and *bad* should be understood as *attributive* terms, i.e. as terms whose meaning cannot be understood as detached from the set of

descriptive features associated with the item that is being evaluated [Geach, P. 1974/1956. p. 67. Foot, P.2002/1958. p. 102, 103-105. 2002/1961. p. 136. 2002/1972. p. 151][23].

The revival of moral realism was connected with three additional consequences. Some of them, or so I will suggest, had an immediate impact on the realocation of the content of *Meta-ethics*. *First,* by recognizing a possible and direct link between evaluations and descriptions, moral realists gave priority to questions related to the epistemology of normative judgments, that is, to issues related to the justification of our evaluative utterances.

Second, by assuming a non-standard view on moral motivation [see Geach, P. 1974/1956. p. 69-70. Foot, P. 2002/1961. p.134-135, 143. 2002/1958-59. p. 122] a set of motivational issues became frequently pursued in those years. Something in between philosophical psychology and theory of motivation emerged as a central topic in *Meta-ethics*. And this happened during or after the revival of moral realism.

Finally, by reflecting on the conditions that basic evaluative uses had to fulfil, moral realists endorsed the hypothesis that the mental state expressed by a moral utterance should be functionally related to a sort of cognition or belief. But besides of this concrete commitment, moral realists implicitly accepted a well-established heuristic principle. In terms of such principle, the meaning of a statement should be determined by asking for the attitude it expresses. As a general result, they defended that *Meta-ethics* should be pursued by focusing on a psychological question about the type of thought underlying our moral utterances.

2.2.2 From an a priori methodology to a broader conception of conceptual analysis

The most important shift in methodology during the period we are concerned now (1930 to 1960) focused on the rejection of the primacy of *a priori model of conceptual analysis*. I will assume here that this methodological reaction was articulated along three different and sometimes interrelated levels. It was articulated either by taking into account the previously noted widening of the content of *Meta-ethics*, or by assimilating a set of advances located at external branches of philosophical inquiry or, finally, by a better understanding of the sort of

---

[23] Although Foot agreed on the emotive, prescriptive or commendatory character of evaluative terms, evaluative realism favoured a distinctive thesis at the semantic level. Realists argued that the grasp of a set of descriptive features linked to the evaluated item is a necessary condition to properly use an evaluative term. They highlighted the centrality of a certain level of evidential accuracy to evaluate any item, rejecting the *sufficiency* of the dynamic or motivational dimension to fix the basis of any genuine evaluation. See Foot, P. 2002/1961, p. 133.

relation to be established between *Meta-ethics*, philosophical methodology, and science. Keeping this caveat in mind - and assuming that I do not have yet the background to offer a detailed account of this complex phenomenon - I will be focused in what follows on a single dimension of this complex process. I am referring to the set of methodological gains derived from the criticism on Moore's methodological model [Darwall, S. et al. 1992. p. 121-122. Timmons, M.1999. p. 25-29].

Remember that Moore aimed at offering *reductive analytic definitions* (RD) of our core moral concepts. In Moore's terms, for reductive naturalism to be plausible a general meta-statement whose general form is $e = N$ (where N stands for a descriptive sentence and e for a sentence including non-descriptive or evaluative terms) should be available. Moore also assumed a central intuition about how such analysis should be unfolded. He believed that any non-moral term or expression contained in the N-side of the definition relating evaluative and non-evaluative properties should express the same thought or proposition, the same *idea*, than the evaluative term contained in the e-side of the definition. Any proposed *natural* term aimed to reduce the meaning of a *moral* term should preserve, so to speak, the simple thought expressed by the moral statement in which the evaluative term was embedded. By using a by now widely established terminology, Moore was committed with *a synonymy test of identity* (ST) [Timmons, M. 1999. p. 27. Sturgeon, N. 2006a. p. 96. Dancy, J. 2006. p. 129][24]. Once RD and ST are in place, Moore's basic goal can be encapsulated by asking a simple question: Can we define a basic moral term (*good* or *right*) in terms of a natural property or set of natural properties without infringing the intuition encapsulated by ST?

Let us suppose that a naturalist is prone to defend that *good* means *pleasure*. Or more accurately: that *good* is identical to *pleasure*. A good answer to the general question about *which things are good* is *whatever thing is pleasant*. Thus, let us assume that the naturalists believes that

(1) good = pleasant

So, in saying

---

[24] Another way to present the point is to establish that $e = N$ if and only if it is analytically true that $e = N$, that is, if e = N is true in virtue of the a priori meaning of e and N. See Harman, G. 1977. p. 19-20. Brink, D. 1989. 163-167 and Smith, M. 1994. p. 28.

(2) x is good

he is affirming, in fact, something like

(3) x is pleasant

Moore famously asked a further question to such naturalist, which focused on the conceptual status of the definition of *good*. Moore asked:

(4) Is the pleasant *good*?

Assuming the identity proposed by (1), (4) should be understood, or so Moore believed, as meaning something close to

(4') Is the pleasant *pleasant*?

But although (4') was clearly implied by (1) and (4), Moore made a philosophical fortune by noting that (4) and (4') are not expressing, against ST, the same proposition. They are not giving voice, to use Moore's own terms, to the same thought or idea. While (4) is open (because it could be false), (4') is giving voice to a tautology. But if it is so, the identity expressed by (1) cannot be true; (1) cannot support the condition assumed by Moore as an essential constraint on any reductive definition - ST. Therefore, the property of *being good* cannot be identical to the property of *being pleasant* [Miller, A. 2003. p. 14].

This argument - *the open question argument* - was attacked almost from the very moment it was formulated[25]. In a certain way, and keeping all the caveats noted at the beginning of this section in mind, one of the most important methodological developments in *Meta-ethics* during those years moved around a critical assessment of the philosophical methodology involved in Moore's argument. Michael Smith has resumed the central case against Moore's

---

[25] Some people simply stated that Moore's argument did not deserve to be called an argument at all. They claimed that it assumed precisely the point that it was trying to establish, that is, that naturalism (in the form expressed by identity's statements like (1)) cannot be true. William Frankena, for example, pointed out that the *openness* required to distinguish (4) from (4') assumed all along that (4) was carrying a different sort of information than those expressed by (4'). By defending that an essential part of the content expressed by (4) was not contained in (4') Moore's argument implicitly accepted the falsity of naturalism before any sort of proof. See Frankena, W. 1974/1939 for this point.

argument in the following terms [see additionally Railton, P. 1990. pp. 155-156. Timmons, M. 1999. Sayre-McCord, G. 2006. p. 50. Sturgeon, N. 2006a. pp. 96-97. Dancy, J. 2006. p. 130] :

> "The question is therefore whether non-subjective naturalists are obliged to support the identity claim in this way. And the answer is that they are not. For, once we recognize the distinction between *a priori* and *a posteriori* necessities, we see that two concepts F and G may allow us to pick out the same property despite the fact that the claim 'X has F' is not analytically equivalent to the claim 'X has G'" Smith, M. 1994. p. 28

The criticisms focused on the meta-philosophical status of Moore's argument originated two interrelated consequences. *First*, they helped to revive a substantive meta-ethical option, which had been severely ignored since Moore's formulation of the open question argument. I am referring to the revival of meta-ethical naturalism during the fifties and sixties. *Second*, the revival of naturalism became a real option only after the semantic conception endorsed by Moore was debunked by an improved semantic account. This fact could help us to grasp an important feature emerging during this period: Some advances coming from other philosophical realms - Quine's attacks on the analytic/synthetic distinction [Quine, W. O. 1951], later accompanied by the emergence of a causal theory of reference [Kripke, S. 1972] and the progressive imposition of functionalism in philosophy of mind [Putnam, H. 1975] - started to be incorporated into meta-ethical discussions. They helped to model a more informed standpoint from which to attack some core meta-ethical questions. And they opened the discipline to philosophers working on external branches of inquiry

2.2.3. From the instrumental dependence between Normative ethics *and* Meta-ethics *to the full independence of* Meta-ethics*.*

Around 1960 the situation of *Meta-ethics* could be characterized along two broad claims:

(a) The subject matter of *Meta-ethics* (what I called its *content*) ceased to be dominated by a linguistic inquiry narrowly understood. A pragmatic turn, psychologically orientated, and the emergence of motivational issues could be located at the centre of *Meta-ethics* at this period.

(b) The classical methodology associated with the program of conceptual analysis lost its support in between 1955 and 1970. This favoured an important widening in *Meta-ethics* beyond of the simple rejection of Moore's argument. A clear and definite feeling of methodological aperture permeated the meta-ethical practice during these years. This aperture favoured the progressive inclusion of conceptual tools coming from the philosophy of mind, epistemology and ontology. *Meta-ethics* started to be connected to other philosophical branches in a systematic way.

Taking both points for granted, I will now sketch another important feature of *Meta-ethics* during this period. I am referring to the high degree of independence reached by *Meta-ethics* during this period. In a certain way, *Meta-ethics* evolved from Moore's *instrumental* role in respect to *Normative Ethics* to a stage where *Meta-ethics* gained a separate or autonomous status from any normative concern.

Take, for example, Ayer's conception[26]. In terms of his semantic criterion of demarcation, an *instrumental* model connecting *Meta-ethics* and *Normative Ethics* faced serious problems from the very beginning. In Ayer's sense, as a natural consequence of the exclusion of our first-order moral statements from the domain of semantic respectability, the instrumental dependence established by Moore was not even an option to connect *Meta-ethics* and *Normative Ethics*. How, Ayer asked, could we establish an instrumental relation between two discursive domains if they are not even governed by the same kind of rules? How could we relate *Meta-ethics* and *Normative Ethics* if the latter domain is falling short in respect to the

---

[26] Ayer accepted the prevalence of an external criterion to constrain the intelligibility of any statements expressed by a meta-ethical theory – see Ayer, A. 2001/1936. p. 109. Roughly, by *intelligibility* Ayer understood a condition had to be applied to any discursive practice to be considered as meaningful. In Ayer's view, this condition was understood as meaning something very close to 'possibility of fixing truth-conditions'. So, a given statement could be understood as cognitively meaningful if and only if the conditions where the statement is true could be determined - as I noted before, his own conception of truth, if fully unfolded, supported a less stringent view about the status of our moral opinions.

semantic criterion of respectability, conforming a pseudo-discourse? By Ayer's own lights[27] we can participate in our moral practice by endorsing certain commitments, but no instrumental connection can be established between a philosophical approach to these practices (*Meta-ethics*) and the practices themselves. The latter simply do not fall under our cognitive domain of competence[28]. *Meta-ethics* should be understood, or so Ayer argued, as a full autonomous domain of inquiry, without any kind of connection with our first-order moral opinions.

This general consensus, however, collapsed years after. As I will argue later, the progressive emergence of some powerful systematizations of our first-order moral opinions - i.e. the emergence of moral theory as a distinctive branch of inquiry [Rawls, J. 1951 and 1971] - motivated the revision of Ayer's view [Frankena, W. 1951]. After Rawls, those normative questions that had been previously ignored became philosophically respectable. It was assumed that we should simply turn our attention toward the normative realm. By doing so we could achieve more definite results than by discussing a set of preliminary questions about the meaning of our moral opinions or about the sort of evidence to be adduced in support of them [Darwall, S. Gibbard, A. Railton. P. 1992. p. 121-122

2.3 Contemporary *Meta-ethics*

Until now, I have presented *Meta-ethics* as evolving along three basic dimensions, *content*, *methodology*, and *scope*. As result of it, by 1970 *Meta-ethics* became to be committed with three basic claims:

- - *Meta-ethics* should be concerned either with a psychological question about our moral opinions or with offering the basis for a general theory of justification of moral opinions *(priority of epistemology and psychology over semantics and ontology)*

---

[27] *Meta-ethics* should face a negative task: to offer a demonstration that ethical concepts are, in fact, pseudo-concepts, concepts whose function is entirely outside of the cognitive realm. Ayer statements in this sense are fairly strong: "We find that ethical philosophy consists simply in saying that ethical concepts are pseudo-concepts and therefore no analysable (…) There cannot be such a thing as ethical science, if by ethical science one means the elaboration of a 'true' system of morals". Ayer, A. 2001/1936. p. 116

[28] Ayer wrote about value judgments: "We shall ourselves to show that in so far as statement of value are significant, they are ordinary, scientific statements; and that in so far as they are not scientific they are not in the literal sense significant, but are simply expressions of emotion which can neither true or false" Ayer, A. 2001/1936. p. 104

- *Meta-ethics* should develop a wider sense of possibilities associated with conceptual analysis beyond of notions such as analyticity or conceptual truth *(broadening conceptual analysis)*

- *Meta-ethics* could be pursued with full independence in respect to normative issues *(independence between Normative Ethics and Meta-ethics)*

This trio of claims was importantly altered after 1970. As before, we can summarize the changes by focusing on the *content*, the *methodology*, and the proper *scope* of Meta-ethics during this period. Let me start by noting some things about the changes on the components conforming the second-order stance.

2.3.1. *The centrality of rationality, explanation, and minimalism in current* Meta-ethics

After 1970 *Meta-ethics* began to be interested in three new topics. Firstly, *Meta-ethics* started to be focused on the content of our *judgments of reasons* and *rationality* from an explanatory and a normative perspective. Secondly, *Meta-ethics* began to be concerned with the status of our *normative explanations.* Some people assumed that by asking some questions about our explanatory practices in the moral domain we could gain some kind of insight into some core meta-ethical questions, i.e. about the semantics, ontology, and epistemology of our moral judgments. Finally, the emergence of *minimalism* offered a new perspective from which meta-ethical expressivism could accomodate some realist platitudes about our moral discourse and practice. Being fully aware of the kind of problems lurking the kind of generalization I have just made about the way *Meta-ethics* was structured from 1970 onwards, let me open this section by pointing to some things about reasons and rationality in recent *Meta-ethics*.

Among those who stressed the centrality of reasons and rationality, Thomas Nagel's groundbreaking work on moral motivation and rationality exemplified the kind of substantive concerns that emerged at this stage[29]. Before Nagel's work, the topic of moral motivation had been widely treated in the domain of moral theory [Falk, W. D. 1948 and Frankena, W. 1958.

---

[29] Thomas Nagel, *The Possibility of Altruism*, Princeton, Princeton University Press, 1970. See Wallace, R. J. 2006/1990 for a good assessment of Nagel's importance.

See Raz, J. 1978. Introduction for an good overview]. However, most of the times those motivational issues had been used either as central premises on a variety of arguments in support of non-cognitivism or along with the quite different problem of the force of our moral obligations. Nagel, on the contrary, attempted to highlight the status of our moral judgments from a different perspective. In a certain sense, he opened a new path for *Meta-ethics* beyond the arguments resting on the motivational dimension of moral judgments.

In essence, Nagel assumed that we could get some important insights about the psychological nature of our moral opinions - about the type of objectivity to be attained by them - if we focused our attention on a preliminary question about the reasons we have in non-practical contexts [Darwall, S. et al. 1992. p. 131][30]. He was concerned with a set of interrelated questions. First, he asked: How can we comply with certain basic requirements of rationality (case of *prudence*)? Once a positive response was offered, he asked a quite interesting question: How can we use this explanation to understand how we are motivated to act morally? [Nagel, T. 1970. p. 2]. Nagel's basic insight assumed that it would be possible to gain some understanding of the nature of our moral motivation by (i) paying attention to the effective workings of our prudential reasons and by (ii) extracting some consequences from these normative structures for the scope, structure, and psychological nature of our moral motives [Nagel, T. 1970. p. 14].

In Nagel's view, once certain structural considerations about our own identity are secured, we can offer an explanation of how we are normally motivated by the prudential reasons we recognize - *even when* such reasons are not related to our current motives. Nagel assumed that once such explanation were in place for *prudential* reasons (and Nagel stressed the fact that the agent effectively takes into account some of his temporally-neutral interests in his actual deliberation about what to do now) we could further explain how our *categorical* moral principles attain such level of compliance. Even when our moral principles are not related to our desires, Nagel famously pointed, we can explain (by attending to the same range of capacities that enable our prudential motivation) why we are disposed to take into account

---

[30] Nagel announced what today is customarily called *moral psychology*. In a recent survey devoted to present the goals and problems of this branch of study, Wallace writes: "Moral psychology is the study of morality in its psychological dimensions. Its unity and interest as subject derive from its connection to the larger subject of moral philosophy, conceived as the study of normative demands on action in general, and moral demands on action in particular (…) If we want to understand morality as a normative phenomenon, we must understand the psychological states and capacities that make it possible for it to regulate people's lives in the way it aspires to" Wallace, R. J. 2005. p. 86. Wallace's formulation is clearly inspired by Nagel.

some agent-neutral interest to determine what to do from a moral perspective [Darwall, S. et al. 1992. p. 135. Wallace, R. J. 2006/1990. p. 22-23].

Nagel extracted two consequences from his case around altruism. *First*, he stressed a substantive consequence. He noted that if the very same psychological framework (a cognitive one) supported the non-temporally relative character associated with our prudential principles and the agent-neutral import of our moral rules, meta-normative cognitivism should be endorsed [Nagel, T. 1970. p. 14 about the notion of *structural considerations* and Chapter X for the treatment of altruism].

Second, Nagel's overall approach announced, as I noted above, a new methodological path for *Meta-ethics*. By following Nagel's insights, several people explored in successive years general issues about normativity with an eye on their meta-ethical consequences. They did so either by endorsing other philosophical views about the nature of reason [Parfit. D. 1984. Gauthier, D. 1986] or by developing themes choose from a broadly kantian perspective [Darwall, S. 1983. Korsgaard, K. M. 1986. 1998. Wallace, R. J. 2006/1990]

But apart from the emergence of issues on reasons and rationality, this period witnessed the birth of a new sort of concern in the second-order stance defining *Meta-ethics*. I am referring to the prevalence of *explanatory* topics to resolve a set of core meta-ethical disputes [Zangwill, N. 2006. p. 264]. In a certain way[31], since Gilbert Harman asked about the *explanatory* status of our moral observations, the generic topic of *moral explanations* has become more and more relevant to understand the current shape of *Meta-ethics* [Harman, G. 1998/1977. Thompson, J. J. 1996. Sturgeon, N. 1998/1985. 2003. 2006 and 2006a Railton, P. 2004/1986. 1990 and Leiter, B. 2001a].

Harman endorsed a general argument aimed to question the objectivity of our moral opinions [Shafer-Landau, R. Cuneo, T. 2007. p. 329]:

> A. We have reasons to believe that moral facts exist only if they explain the formation of our moral judgments

---

[31] Surely, explanatory topics are so pervasive that we cannot reduce them to a simple argument or debate. Classical debates around moral motivation, for instance, could be constructed as instances of a general explanatory requirement focused on the ambition of the meta-ethical theories that assume the existence of moral facts and moral beliefs. In terms of this particular, moral facts do not exist because they cannot explain why we are motivated to act in accordance with our moral opinions. Here, however, I am narrowing the explanatory constraint, focusing it on the explanatory plausibility of moral facts and moral beliefs to explain our moral observation. See Shafer-Landau, R. Cuneo, T. 2007. p. 328

B. Moral facts do not explain the formation of our moral judgments

Therefore, we do not have reasons to believe that moral facts exist

He supported the previous argument at the level of B by appealing to a well-known argument. I will offer here a brief reconstruction of it:

(1) We are able to respond to some perceived phenomenon by making an observation - *factual premise*

(2) An *observation* is an assertion not constitutively dependent upon any previous process of reasoning - *conceptual premise*

(3) *There goes a proton* and *pouring gasoline on a cat is wrong* are instances of physical and moral observations respectively

(4) Any observation about p^{32} endorsed by A - (O_A) - can count as confirming evidence for the existence of p to the extent that it is reasonable to explain O_A by assuming that p exists - *inference to the best explanation* [Harman, G. 1998/1977. p. 88]

(5) In the scientific case - *there goes a proton* - we can explain the agent's *physical* observation only if we suppose that it is really the case that *there is a proton*

(6) In the moral case - *pouring gasoline on a cat is wrong* - we can explain the agent's *moral* observation *without* supposing that it is really the case that poring gasoline on a cat is wrong

Therefore, by (4) and (5), an agent's *physical* observation about *protons* offers evidence about the existence of *protons*

Therefore, by (4) and (6), an agent's moral observation about the *wrongness* of a certain action does *not* offer evidence about the existence or reality of a moral property associated with a certain act-type (*poring gasoline on a cat*)

---

[32] Where *p* stands either for a physical object - *proton* - or for a moral property - *wrongness*.

Harman's conclusion has been the focus of a still ongoing debate between (a) those philosophers who defend an sceptic view about the ontological status of our moral concepts based on Harman's basic insights, (b) those who accept the validity of the explanatory intuition underlying Harman's case without endorsing his negative conclusion about the status of moral properties (Sturgeon, Railton) and, finally, (c) those who favour a non-explanatory approach (a purely normative approach) to determine the reality of our moral concepts (Hampton).

Although the case around moral explanations is not close yet, I tend to believe that those located at (b) have presented a more sophisticated array of arguments than those favouring (a) or (c). Some philosophers have recently defended, in this vein, that even if moral facts are not real in Harman's sense, they are helping to *structure* our commonsensical moral explanation by *programming* them. It is because of this *programming* role, they have argued, that *moral facts* (although not causally relevant in the sense that natural facts are) should be understood as real [Cuneo, T. 2006. Jackson, F. Pettit, P. 1990]. The ball, or so it seems to me, is now on the sceptic's roof.

But despite of the importance of *rationality* and *explanation* in contemporary meta-ethical theorizing, I tend to believe that the most important development in recent *Meta-ethics* has moved around the progressive centrality of *minimalism* to sustain a certain variety of meta-ethical theory [Horwich, P. 1990. Wright, C. 1992].

The importance of *minimalism* to understand the shape of recent *Meta-ethics* cannot be fully grasped unless we keep in mind how certain platitudes about our moral stance (chapter 1, 4.1 and 4.2) fix the *focus* of *Meta-ethics*. It is because of the fact that *minimalism* has offered a better framework for *expressivism* to *accommodate* certain core platitudes about our moral practice (see chapter 1, section 3), that *minimalism* is so important to understand the shape of current *Meta-ethics* [Smith, M. 1994a. Divers, J. Miller, A. 1994. Dreier, J. 2004].

In the terms just assumed, the way in which *minimalism* has altered the meta-ethical debate is pretty clear. To put it simple, *minimalism* has facilitated *expressivism* with the philosophical tools required to accomodate some central features about our moral language, ones that have been traditionally appealed to in support of moral realism. Thus, while *classical* expressivists (Ayer, Stevenson, and Hare) offered a revisionary account of the semantics of our moral language – not being concerned at all with accomodating the realistic grammar of our moral language into an expressivist semantic mould – *minimalism* has offered

– at least the one endorsed by some *contemporary* expressivists (Blackburn, Gibbard) - the possibility to accomodate some realistic features of our moral language into a *non-factualist* account of the semantics of our moral discourse [Horwich, P. 1993. p. 143. Blackburn, S. 1998. p. 75-83. Gibbard, A. 2003. p. x. Preface. Van Roojen, M. 2004. p. 30. Sinclair, N. 2007. p. 345].

As I suggested in the previous chapter, some people tend to think that the assertoric contexts where our moral opinions sometimes appear would secure by themselves a commitment with a set of heavy-duty metaphysical concepts - moral *truth*, moral *belief* and moral *facts* (Chapter 1, section 4.1.1). They assume that as far as our *moral* sentences possess a number of features thought to be characteristic of those sentences "wearing propositional clothing" [Sinclair, N. 2007] we would be entitled to speak of *truth*, *facts*, and *beliefs* also in relation with our moral opinions. Against this set of substantive commitments, *minimalism* tries to reasses the metaphysical significance of such commitments by deflating them. In terms of minimalism about *truth*, for instance, the proper way to understand truth-ascriptions is to ask about the dynamic of the speech-acts where the truth-predicate is used [Dreier, J. 2004. p. 24]. When we focus on the pragmatic nature of such acts (and not on the property that the truth-predicate supposedly denote) we find out that ordinary uses of *truth* opperate in accordance with the following schema:

M **TRUTH** 'S' is true if and only if S

(where 'S' stands for a sentence in a language). In terms of the previous schema, when one is using the truth-predicate in relation with S, what one is doing is simply affirming S. Our truth-predicate normally behaves as a meta-linguistic device aimed to endorse a sentence or a set of sentences in a variety of discursive contexts. If minimalism is right, then, nothing about the typical contexts where our truth-predicate appears demands a realist framework to explain those linguistic practices involved in such appareances. The schema above offers by itself a neccesary and sufficient condition to grasp all we need to know about truth in *any* context [Horwich, P. 1990. p. 18-23].

Accordingly, if faced with an utterance such as

(1) It is true that torturing babies is wrong

expressivists should be able to explain the sense in which (1) is true by simply applying the previous schema to (1):

(1') 'Torturing babies is wrong' is true if and only if torturing babies is wrong

(1') would tell us all we need to know about (1) being true[33]. In this particular case, by uttering (1) the speaker is endorsing himself a negative attitude toward torturing babies [Blackburn, S. 1998. p. 79. Ridge, M. 2007. p. 85].

But *minimalism* does not stop here. The same deflationist style referred above can be also applied to moral *facts* and moral *beliefs* [Rosen, G. 1998. p. 396. Gibbard, A. 2004. p. 182-183. Dreier, J. 2004. p. 26. Cuneo, T. 2007. Chapter 6]. In the case of any type of *fact* – and note how the notion of *fact* is intimately connected with the overall semantic picture favoured by realism – *minimalism* would assume the following equivalence:

M ꜰᴀᴄᴛꜱ *that p* is a fact if and only if *p*

In the case of any type of representational psychological state minimalism would propose the following formula to debunk a certain metaphysical reading of the status of our beliefs:

M ʙᴇʟɪᴇꜰꜱ *p* represents *that p* if and only if p - and p is subject to certain standards of use

Thus, the simple fact that p can be properly asserted - being able to appear in contexts of logical embedding, propositional-attitude ascription, and truth-predication - is a neccesary and sufficient condition to secure the representational status of p. Nothing more is required to assume that when we say that torturing babies is wrong we are expressing a belief (the belief that torturing babies is wrong) that to accept that (i) we are able to assert that 'Torturing

---

[33] *Minimalism* offers, in fact, a two-level account of the sense in which the truth-conditions of (1) are fully unfolded by (1'). *Minimalism* starts by endorsing a certain view about *truth-aptness* or *truth-assessability* It is assumed by *minimalism* in this sense that for any sentence (S) located at the left-hand side of (1')-schemas, S is *truth-apt* only if (a) S belong to a syntactically demarcated category - logical embedding, belief-ascription, etc. - and (b) S is respecting appropriate standards of use (Wright, C. 1992. p. 29). Once assumed the previous story about *truth-aptness*, the *truth-conditions* of S are fixed by (1'). As Smith notes: "Accordingly to the minimalist, a truth-assessable sentence, s, is true if and only if the conditions on s's proper use are met: that is, if and only if things are as s says they are. And to be prepared to say that things are as s says they are we, theorists, need simply be prepared to utter s itself". Smith, M. 2006/1994. p. 424. See Jackson, F. Oppy, G. Smith, M. 1994 for a different story about *truth-aptness*

babies is wrong' and that (ii) such sentence can be embedded in typical logical patterns, subject to truth-predication and so on [Blackburn, S. 1998. p. 79-80. Sinclair, N. 2007. p. 344].

Once minimalism is unfolded, however, its debunking role in minimizing the apparently heavy-duty metaphysical status of certain commitments incorporated into our moral discourse is clear. It is because expressivists have recently appealed to *minimalism* that they are able to accommodate notions such as moral *truth*, moral *facts*, and even moral *beliefs[34]*. Thus, contrary to what *classical* expressivists claimed, *contemporary* expressivism has no urge to minimize certain phenomenna ocurring at the level of the surface grammar of our moral language. They can speak of moral *facts*, moral *truth* and moral *beliefs* without assuming the realist image. They can mimic, so to speak, the realist core commitments without endorsing a realist, metaphysically strong, explanation of such commitments [Dreier, J. 2004. p. 25-26. Horwich, P. 1993. p. 143. Rosen, G. 1998 p. 387]

2.3.2 The emergence of reflective equilibrium and the pervasiveness of naturalism as a methodological constraint

Contemporary *Meta-ethics* can be characterized by its commitment with a certain model of justification (*reflective equilibrium*) and by its acceptance of *naturalism*. When combined, both intuitions depict an overall image of how *Meta-ethics* should proceed, from a methodological level*,* to offer an accurate reconstruction of our first-order moral stance[35]. Let me start by the importance of the ideal of *reflective equilibrium* to understand the shape of our current meta-ethical theorizing.

By *reflective equilibrium* we usually understand, at least since Rawls's groundbreaking work, an epistemic approach to justify a particular moral belief by locating it into wider structure. Rawls distinguished two senses of equilibrium. We could justify a moral belief by

---

[34] Although see Timmons, M. 1999 and Timmons, M. Horgan, T. 2006. In these pieces they accept the minimalist story about truth, developing a different solution to accommodate our talk about moral beliefs. For more about this see Chapter 8, section 4.1

[35] As in the case of the substantive concerns endorsed by meta-ethics, the methodological advances cannot be fully spelled out without a great level of simplification. I am fully aware that several methodological points are necessary being ignored by focusing on the concept of *equilibrium*. For example, a very important topic in contemporary moral theory, which has surely permeated recent meta-ethical approaches to some extent, are the *ambitions* associated with moral theory - when 'moral theory' is understood along some overall normative frameworks such as utilitarianism or deontological approaches. See Williams, B. 1976. 1985. 1996 and Railton, P. 2004 for a recent discussion of Williams's endorsement of *anti-theory* for future reconstructions of *Meta-ethics*.

locating it into a set containing several moral principles, and by trying to produce coherence between our moral beliefs and our principles (*narrow reflective equilibrium*). But we could also justify a first-order moral belief by locating it into a set composed by moral principles and by some basic insights coming from external disciplines (background theories) such as psychology or sociology (*wide reflective equilibrium*) [Rawls, J. 1971. p. 49. See Daniels, N. 1979 for a complete presentation of Rawls's program]

Although it initially focused on the justification of first-order moral beliefs, Rawls's intuition has been recently endorsed in a slightly different sense. It has been suggested, in essence, that it would be possible to find a methodological place for the concept of *reflective equilibrium* when we evaluate the accuracy of the *accommodation* offered by our second-order theories about our moral experience [Daniels, N. 1979. p. 271. Gibbard, A. 1990. p. 25. Timmons, M. 1999. p. 11-12 and De Paul, M. 2006. p. 598-560]. Just as the reflective model can work when it is focused on the justification of our first-order moral intuitions (by making explicit the extent to which such intuitions can cohere with our moral principles and background theories), some people tend to assume that the very same model could guide the process of rationally by choosing among alternative meta-ethical theories. They note, in essence, that the preferred meta-ethical theory should exemplify the maximal degree of *equilibrium* possible. That is, they argue that the preferred meta-ethical theory should represent a maximum degree of coherence (wide equilibrium), understanding *coherence* as a certain degree of explanatory integration between the intuitions that shape our moral stance (see Chapter 1), a set of philosophical thesis about our moral perspective, and a general commitment with a naturalistic view of the world [Daniels, N. 1979. p. 273-274 and 282. Darwall, S. et al.1992. p. 125-126]

But, what do we mean by *naturalism* in this context? For some people, *naturalism* is mainly a metaphysical or substantive thesis. It assumes that any entity, relation, or property (simple or complex) in a given discursive domain can be understood as an element of the natural world, i.e. as a natural property constituted by the kind of entities or relations that natural sciences study [Railton, P. 2004/1986. especially p. 3-4 and 2004/1986a. p. 62-63. 1990. p. 155. Timmons, M. 1999. p. 14. Copp, D. 2007. p. 10. 2007/2003. p. 33, 47-53 and Hornsby, J. 1997. Introduction][36].

---

[36] Naturalism in this substantive sense could be interpreted further as a reductive thesis claiming that, at the very the end, *a particular* scientific stance must be privileged to identify the relevant properties to which *naturalism* points at. Understood along these lines, naturalism is committed with *physicalism*. See Jackson, F. 1998. p 6-8.

But *naturalism* is also a *methodological* ideal that indicates how to proceed when doing philosophy. In this further sense, naturalism defends that:

> "(…) philosophy does not posses a distinctive, a priori methodology able to yield substantive truths that, in principle, are not subject to any sort of empirical test (….) a methodological naturalist believes that philosophy should proceed a posteriori, in tandem with - perhaps a particularly abstract general form of –the broadly empirical science. Railton, P. 1990. p. 156

Naturalism, at least in the *methodological* sense[37], represents a widely assumed intuition in recent *Meta-ethics*. Due to its prevalence, this approach has favoured two general consequences. Firstly, it has helped to establish by itself a wider level of assessment to be met by any meta-ethical theory. It has conformed, to put in some way, a wider level of assessment to be met by any meta-ethical theory. In a certain sense, by assuming naturalism we could be able to pick a meta-ethical proposal from among different, albeit equally plausible, meta-ethical theories.

But besides this external or regulative role, *naturalism* has become central in an additional, internally rooted sense. This general intuition has helped to articulate the influence of a set of empirical branches from which we can extract a vast amount of evidence to support or to revise some of our most common intuitions about the *shape* of our moral stance[38]. It is against the general framework facilitated by *naturalism* that we can understand some recent proposals coming from social and developmental psychology, comparative psychology, and what has become to be known as experimental philosophy. As I take them, these recent branches of

---

[37] *Substantive* and *methodological* senses of *naturalism* are sometimes put apart. We could endorse, for instance, a naturalistic-rooted account of the content of our moral claims by departing from a process of *a priori* analysis. In this sense, we could be substantive naturalists without endorsing naturalism at the methodological level. Inversely, we could embrace a *methodological program* of research committed with naturalism without assuming, at the very end, a naturalist account of the *content* of our moral opinions – because, as expressivism notes, there is no such content to be analyzed. Although the distinction in terms of substance and method is quite pervasive, and surely it allows for the options just noted, I will assume here that, at least in recent discussions, the methodological sense of naturalism has been prevalent. See Railton, P. 1990. p. 156

[38] An important methodological achievement that occurred during this period stresses the importance of a more liberal conception of conceptual analysis. See Stich, S. Weinberg, J. 2001. Stich, S. Doris, J. 2005 and Sosa, E. 2006.

inquiry are trying to implement *naturalism*, so to speak, at the different levels of moral theorizing [Green J. Haidt, J. 2002. Nichols, S. 2004. Joyce, R. 2006. Prinz, J. 2007].

To sum up, contemporary *Meta-ethics* could be characterized, at least from a purely methodological standpoint, by its commitment with a model of justification around the concept of *reflective equilibrium* and by a general endorsement of *naturalism*. As a result of this double commitment, *Meta-ethics* has witnessed the progressive emergence of empirical approaches and the revision of the type of *accommodation* that should guide the construction of any second-order theory about morality. If I am right, one of the most definite advices to be extracted from contemporary *Meta-ethics* is simply that our second-order theories of morality should attempt a *wide* level of accommodation for our moral intuitions, which should take into account results from different branches of inquiry such as psychology, anthropology, and sociology.

2.3.3 Achieving a high degree of inter-dependence between normative and meta-ethical issues

I noted before (section 2.2.3) that the golden age of *Meta-ethics* witnessed a lack of interest in normative issues. *Meta-ethics* and *Normative Ethics* were pursued, or so I suggested before, along separate routes. Contemporary theorizing about morality has reconsidered this independence. As result of it, a strong sense of *interdependence* between *Meta-ethics* and *Normative Ethics* has emerged. This interconnection between first-order and second-order issues could be explained by appealing to certain social or historical factors [Darwall, S. et. al. 1992]. Or it could also be explained by simply appealing to the ambitious scope of some landmark pieces of work in recent moral theory. To some philosophers, a good route to overcome the *impasse* in *Meta-ethics* would consist in attacking normative and second-order issues on a par. If I am right, a contextually dependent factor (*impasse*) could help to explain why *Meta-ethics* has become to be pursued along with normative or substantive issues (more about this in the next section).

Remember the intermediate view I depicted above (2.3.3). In terms of it, *Meta-ethics* cannot be connected (not even instrumentally, as Moore defended) to *Normative Ethics* because of the lack of semantic credentials of the normative, firs-order, discourse. Although it was a radical view, the positivistic framework did not hold a long period of philosophical prevalence. From 1960 to 1970, a reconsideration of the relationship between *Normative*

Ethics and *Meta-ethics* was achieved. As I noted above, as a direct consequence of a group of pressures coming from outside moral philosophy and the work of certain major figures (the impact of Rawls's groundbreaking *Theory of Justice* is commonly cited at this stage), second-order issues about morality became related to substantive topics in normative theory. Once we assume this framework we can ask how has the process of interconnection between *Normative Ethics* and *Meta-ethics* has evolved in recent years. Although there are several ways to illustrate the progressive interdependence between these two issues, the most promising routes run along the following lines:

- Some analyses of the concepts of *reason* and *rationality* have helped to establish some general constraints to be applied for selecting a good first-order normative theory [Parfit, D. 1984. Gauthier, D. 1986. Scanlon, T. 1998].

- Some first-order accounts of *autonomy* have recently shifted toward second-order analyses of *valuing* and the capacities that enable such practice [Korsgaard, C. 1998. Wallace, R. J. 2006/2000 and 2005. Bratman, M. 2006. Introduction. Watson, G. 2004. Frankfurt, H. 2005 and 2006]

- Finally, some second-order analyses of the role and status of principles in our normative thought have questioned the plausibility of certain first-order moral theories [Dancy, J. 2004. Ridge, M. McKeever, S. 2006. Lance, M. Little, M. O. 2006 and 2006a].

As the above lines of inquiry suggest, there is an intimate and fruitful relation between normative and second-order issues in the contemporary theorizing about morality. Thus, although *Meta-ethics* has evolved toward a quite specific domain of inquiry, many people accept that paying attention to the nuances embedded into our first-order system could help us to understand the status of morality.

3. The *impasse* in current *Meta-ethics*

I have devoted the previous section to expose what I take to be some of the main lines of development of *Meta-ethics* during the last century. Evidently, this historical overview has been conceived along a partisan line of interpretation. It expresses, to say the less, an idiosyncratic and partial view of the historical configuration of *Meta-ethics*. To some people, some important lines of development should be included into my account while others should be removed from it. In any case, and assuming in advance the legitimacy of several concerns that may be adduced against my personal reconstruction, the previous description has been mainly dictated by an *hermeneutical* intuition. In brief: I am convinced that by depicting an historical overview of *Meta-ethics* we could understand better a scenario exemplified by current *Meta-ethics*. If I am right the historical account I have just offered is useful only insofar as it helps us to understand the plausibility of a certain scenario. But which scenario is it?

Remember that in the previous pages I have assumed that besides the classical philosophical concerns about morality (*semantics*, *ontology*, and *metaphysics*), *Meta-ethics* is actually concerned with a set of additional questions about our moral stance - the status of reasons and rationality, the nature of moral explanations, the psychological import of our normative thought, methodological assumptions about the nature of the conceptual analysis, the status of our actions, general issues on theory of normativity, and some concerns related to the credentials of naturalism. As I have argued in the previous section, all these issues have contributed to conform the current complex shape of *Meta-ethics*. Clearly, the shape of Meta-ethics I have favoured here is far away from the *packed-and-ready definition* of *Meta-ethics* contained in many introductions to moral philosophy. *Meta-ethics* is dealing, of course, with moral opinions linguistically framed. If I am right, nevertheless, current *Meta-ethics* locates such peculiar acts of speech into a wider net of regularities and patterns, trying to *accommodate* such complex phenomenon (our moral stance) into a comprehensive, second-order, explanation. This complex view can be encapsulated by means of the following schema:
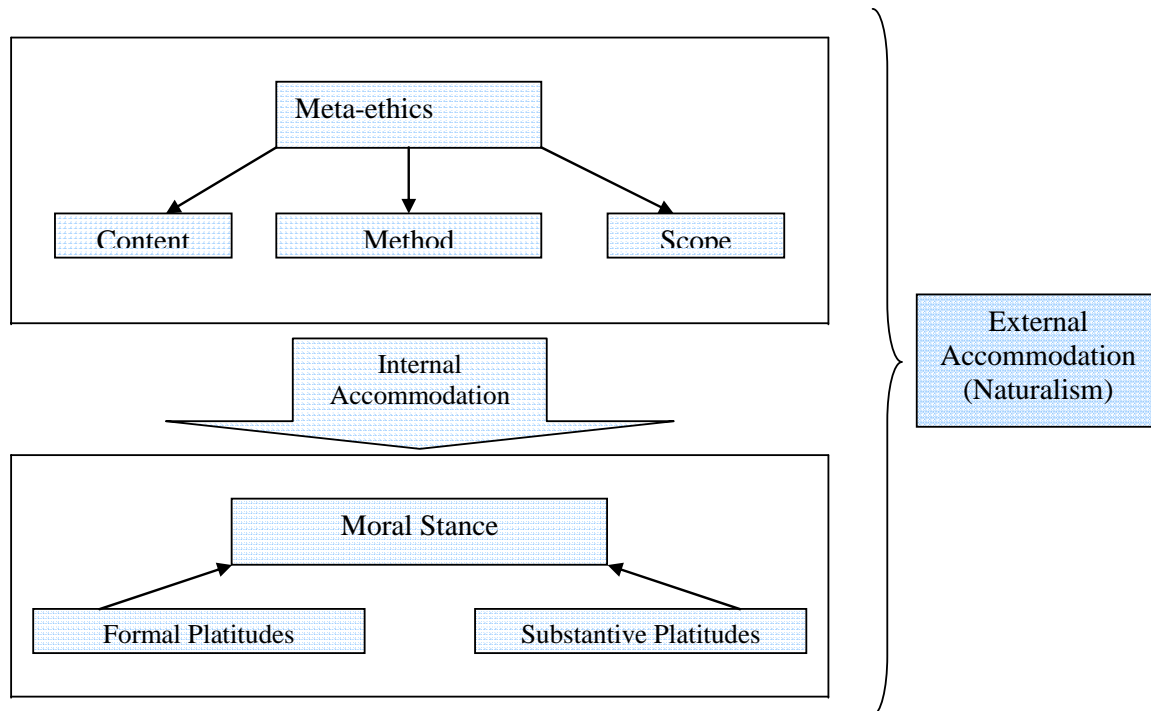
Keeping in mind the schema above, the image of *Meta-ethics* that I tried to motivate in the current chapter announces the possibility of a certain scenario. Here, two meta-ethical theories would aspire to a similar degree of *accommodation* (or equilibrium) in respect to the platitudes defining our moral stance. Both theories would offer, so to speak, two theoretical reconstructions of a common domain of first-order intuitions (using the conceptual tools proper to *semantics*, *ontology*, *epistemology*, *moral psychology*, *theory of explanation*, and *theory of rationality*). And both theories would respect the external constraint posited by *naturalism*. To put otherwise: they would explain the status and functions of our moral utterances by resting on the substantive and methodological assumptions endorsed by a naturalistic view of the world and by assuming the same set of intuitions about our moral stance. In a situation like this both theories are exemplifying an *impasse*[39].

---

[39] Shafer-Landau unfolds, at a certain stage, something close to the general framework I am presupposing here. Speaking of how issues of moral disagreement could affect the prospects of moral realism he notes: "Whether the realist's account is satisfactory on what else the realist has to say – about ontology, epistemology, moral motivation, and reasons for action. And it depends on what the anti-realist has to say about these things as well. The ultimate question must be focused on which picture offers the best comprehensive meta-ethical view, and even if there were some advantages gained by anti-realism on matters of disagreement, that advantage might be overridden when surveying the liabilities of realism's competitors and noting realism's strengths. We can't confidently pronounce on the degree to which realisms is weakened by considerations of moral disagreement until we have in hand a well-grounded comparative assessment of how realisms and its competitors face on a whole range of additional issues" Shafer-Landau, R. 2003. p. 219.

To clarify further the scenario that I have in mind, let us return to the set of platitudes that shape the image that we share about our moral stance (Chapter 1, sections 4.1 and 4.2). These platitudes are paramount because of the central role that they play in constituting our moral perspective. *Formal* platitudes about moral *truth*, moral *objectivity*, moral *knowledge*, *explanatory import*, *normativity*, and *action-guidance* would surely play an important role in defining our conception of a moral judgment. And the same goes for some *substantive* platitudes about our moral judgments. All these platitudes are essential to characterize our reflective image about morality. Assuming their centrality then, by saying that *Meta-ethics* is facing an *impasse* we could imply:

- **(1A)** A scenario where two opposite meta-ethical theories (two theories endorsing opposite views about the *semantics*, *ontology*, *epistemology*, *explanatory import*, etc…) are able to accommodate a similar but *not* identical set of platitudes about our moral stance. This scenario exemplifies a *non-strict impasse*.
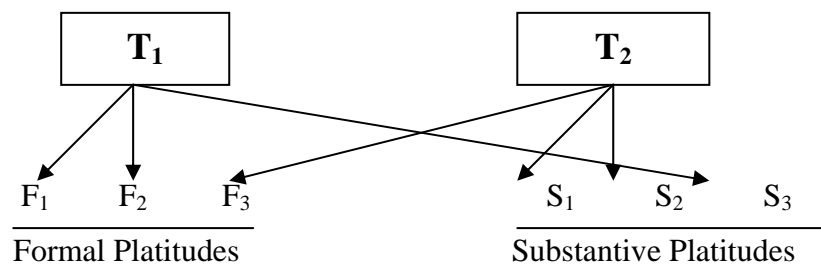
- **(1B)** A scenario where two opposite meta-ethical theories are able to accommodate an *identical* set of platitudes about our moral stance. In my terminology, this scenario exemplifies a *strict impasse*.
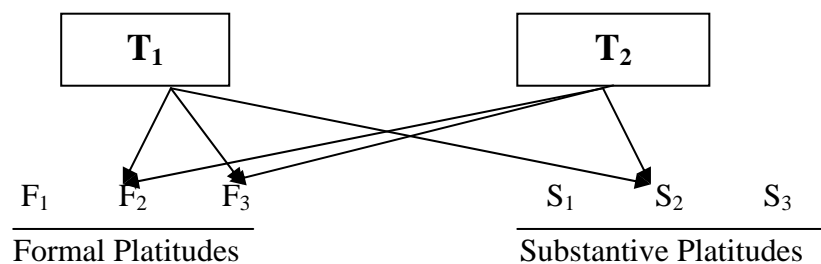
The type of *impasse that* I have in mind is a *strict* one. However, even if the *strict* interpretation is assumed, two additional partitions will emerge as result of reflecting on the *content* of the second-order stance from which we accommodate the set of platitudes that conform our moral stance, and on the role of *naturalism* in the overall process of accommodation. In terms of these additional partitions we could distinguish between:

- **(2A)** *A severe impasse*: **(1B)** plus the fact that the two meta-ethical theories that reconstruct our moral stance are assuming an *identical* set of second-order disciplines (semantics, epistemology, ontology, theory of explanation, etc...)

$$D_1 \quad\quad D_2 \quad\quad D_3 \quad\quad D_4$$

$$\boxed{T_1} \quad\quad\quad\quad \boxed{T_2}$$

$$F_1 \quad F_2 \quad F_3 \quad\quad\quad S_1 \quad S_2 \quad S_3$$

Formal Platitudes   Substantive Platitudes
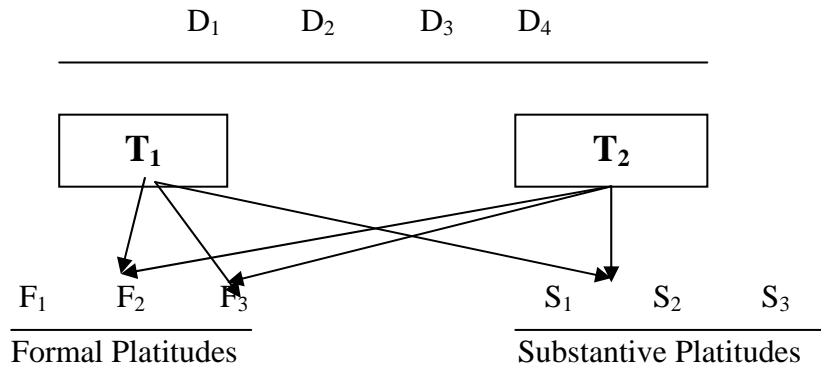
and

- **(2B)** *A non-severe impasse*: **(1B)** plus the fact that the two meta-ethical theories that reconstruct our moral stance are assuming a *non-identical* set of second-order disciplines

$$D_1 \quad\quad D_4 \quad\quad\quad D_2 \quad\quad D_3$$

$$\boxed{T_1} \quad\quad\quad\quad \boxed{T_2}$$

$$F_1 \quad F_2 \quad F_3 \quad\quad\quad S_1 \quad S_2 \quad S_3$$

Formal Platitudes   Substantive Platitudes
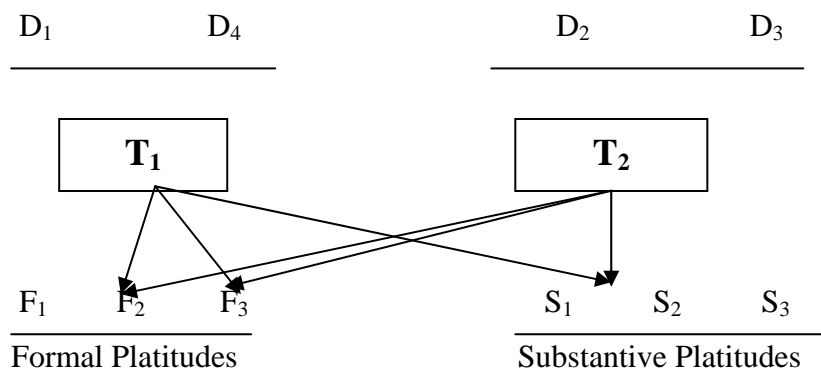
When I speak about an impasse in *Meta-ethics,* I have in mind (2A). However, my position on this issue can be further supplemented by accepting a third additional partition to characterize the shape and structure of the *impasse*, one distinguishing between

- **(3A)** *A full impasse*: **(1B)** and **(2A)** plus the fact that the two meta-ethical theories accept *naturalism*
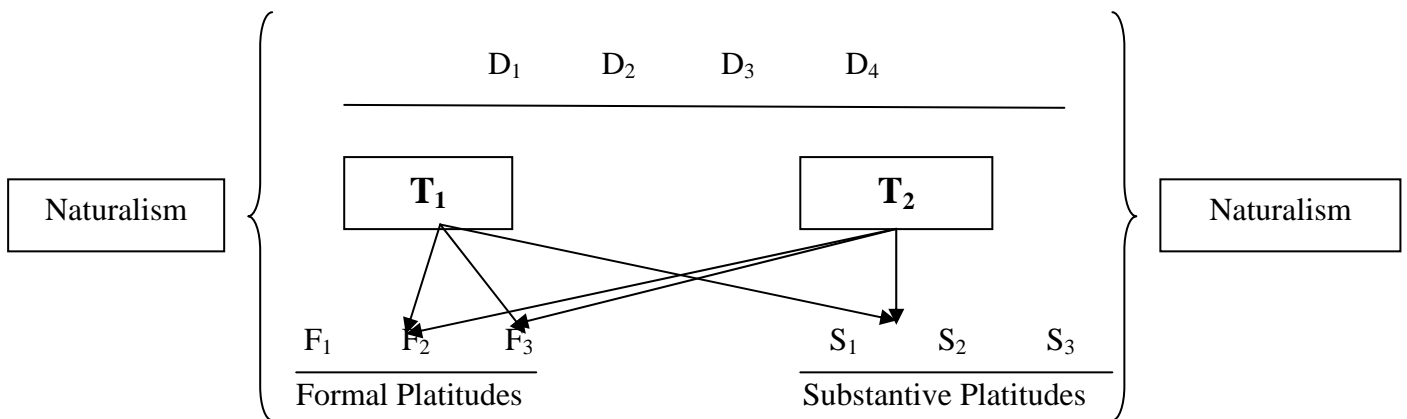
and

- **(3B)** *A non-full impasse*: **(1B)** and **(2A)** plus the fact that one among the theories involved in the *impasse* is rejecting *naturalism*
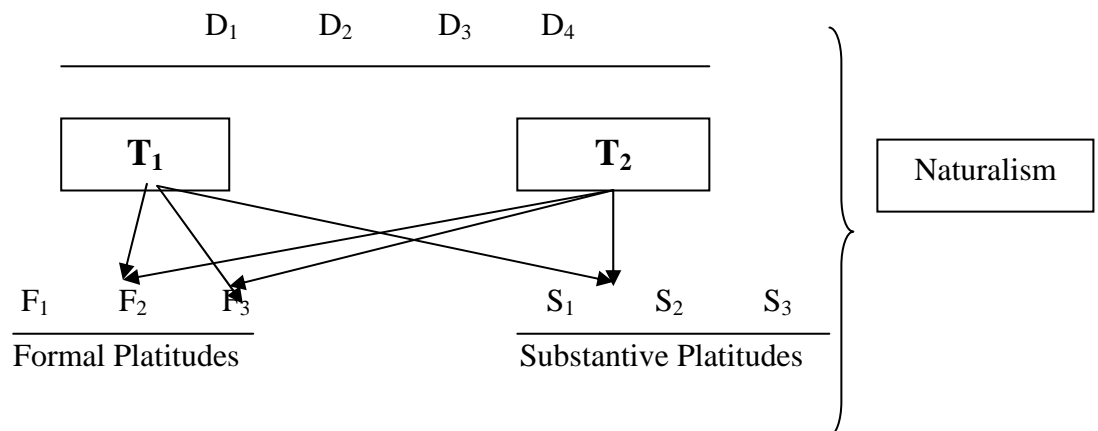
In terms of this last partition, **(3A)** is grasping what I consider a basic ingredient of the deep *impasse* underlying *Meta-ethics*. Thus I am disposed to assume a situation as the depicted by **(3A)** as a good way to describe the current shape of *Meta-ethics*. If I am right the two most prominent meta-ethical theories (*cognitivism* and *non-cognitivism*) could offer two plausible albeit opposite second-order reconstructions of a similar domain of intuitions about our moral perspective[40]. Both theories could appeal to the same set of second-order disciplines in accommodating these intuitions. And finally, both theories could respect the external constraint on accommodation imposed by *naturalism*.

Evidently, we can accept the methodological possibility of (3A) for current *Meta-ethics* only if we presuppose the prominent role that certain semantic developments (*minimalism*) have played in re-defining the terms of recent meta-ethical disputes (see section 2.3.1 this chapter)[41]. It is because *expressivism* can easily accomodate certain core platitudes related to the objective import of our moral opinions that we could understand a situation as the one depicted by (3A) [Dreier, J. 2004. p. 25]. In such scenario, both cognitivists and non-cognitivists would be able to explain the objective import of our moral claims (F3), the action-guidingness of our moral opinions (F2), and the type of mormative import of our moral assertions (S1) (see Chapter 1, section 5). Thus, although both meta-ethical views could agree on the core features about our moral perspective to be explained - accomodating them by referring to the same philosophical categories (*knowledge*, *belief*, *truth*, etc.) – the referred meta-ethical views would explain them by departing from two opposite views about the ontology, semantics, and epistemology of our moral stance. It is in this precise sense in which I am assuming that current *Meta-ethics* is facing an *impasse*.

Surely, if an scenario in line with (3A) was real and not only plausible or methodologically conceivable - if *cognitivism* and non-*cognitivism* exemplified, as a matter of fact, an instance of (3A) - we would *not* be entitled to choose a second-order theory with the feeling that such theory offers a sound and undisputed explanation of the status of our moral practices. As a

---

[40] McNaughton denies it when he writes: "It has already emerged that in judging between our rival meta-ethical theories we need to take into account two things: first, their degree of cohesion with other theories in other areas of philosophy; second, the extent to which they fit moral phenomenology. On the first count both theories establish a reasonable prima facie case. A final decision will have to await more detailed examination of those areas, such as the explanation of action and the relation between science and reality (…) On the second count, however, the realist has strong grounds for claiming that his theory is in much better accord with our moral experience and practice than the non-cognitivist's. McNaughton, D. 1988. p. 64. At page 13 he offers a clear description of the sense in which realism and irrealism offer different views about the status of our moral opinions.
[41] Although see Smith, M. 2004/1998 for an additional construal of the impasse between cognitivists and non-cognitivists. See especially at page 378-379

result of our awareness of such scenario we should rather assume that our preferred meta-ethical choice is, to put it some way, as good as the opposite view. But we should not assume that our preferred choice is better or more justified than the opposite alternative. And of course, we would not be entitled to assume that our preferred theory is true in a substantive sense. As Gideon Rosen notes:

> "At the end of the day we have rather a pair of equally
> legitimate representations of our thought in the area, with
> no clear basis for saying that either is more revelatory of
> its nature than the other" Rosen, G. 1998. p. 401

In fact, some people have suggested that once non-cognitivism is construed around the metaphysical austerity of *minimalism*, the same distinction between cognitivism and non-cognitivism becomes blurred [Gibbard, A. 183-184. Van Roojen, M. 2004. p. 36. Dreier, J. 2004]. They argue, in a nutshell, that there is no way to distinguish a cognitivist from a non-cognitivist if both meta-ethical theories endorse the very same set of core platitudes about our moral stance. Although I will not endorse this variety of *quetism* (see Chapter 8, section 4.1 and 4.2), it should be clear that such meta-philosophical option is grounded, again, on the conceivability of a scenario in line with 3A.

But surely my focus on certain cases where two meta-ethical views are trapped into a methodological *impasse* is not new. I am aware of at least one general argument where an *impasse* between opposite meta-ethical theories is referred as a possibility. David Brink presented an instance of such argument some time ago. It went as follows:

(1) Practical differences[42] are able to determine the explanatory plausibility (or truth)[43] associated with any meta-ethical view.

---

[42] In a clear sense, Brink understands by *practical differences* a level of agreement or equilibrium between the second-order reconstruction offered by a moral theory and the first-level set of commonsensical features usually associated with our phenomenological and inferential commitments in respect to the shape of our moral stance. See Brink, D. O. 1989. p. 82. In another sense, Brink uses *practical differences* in a more direct way, that is, as pointing to certain practical consequences associated with the acceptance of a given meta-ethical theory. Brink, D. O. 1989. p. 83. Although this apparent divergence would not seem to be important by itself, the fact is that, once the desired conclusion is reached, the overall plausibility of it will importantly rest on whether we are concentrating on a sense or another.

[43] Brink's argument is intended in terms of truth while my reconstruction, in line with the content of this chapter, is focused on explanatory or methodological issues. Brink, D. O. 1989. p. 82.

(2) If moral realism purports to be not only intuitively attractive but also explanatory plausible, then it should be able to make overall sense of the practical differences underlying our moral stance.

(3) Both, moral realism and moral irrealism are *equivalent* in terms of how they accommodate those practical differences [Brink, D. O. 1989. p. 82]

Therefore,

- Either moral realism cannot be taken by itself to be an explanatory plausible meta-ethical theory (by 2 and 3) – *argument against moral realism.*

- Or both, moral realism and moral irrealism can be taken as equally plausible from an explanatory standpoint – *argument for the impasse.*

- Or, finally, *practical differences*[44] cannot determine by themselves the explanatory plausibility (or truth) of any moral theory [Brink, D. O. 1989. p. 82-83].

Although Brink's argument was aimed at showing the force of the first conclusion, the second conclusion appears far more plausible to me than the classical non-cognitivist's conclusion. Thus, as far as we are disposed to accept the relevance of *accommodation* (what Brink denominates the level focused on *practical differences*) to establish the epistemic status of any proposed meta-ethical theory, we should assume if not the effective reality of an *impasse*, at least its *possibility* or *conceivability*.

In what follows, I will use on the possibility of an *impasse* inside *Meta-ethics* to argue for an alternative route to do *Meta-ethics*. If I am right, the high degree of sophistication in recent meta-ethical theorizing (the way in which opposite meta-ethical theories can accommodate the very same intuitions about our moral stance) announces the scenario depicted by **(3A).** Surely, the best option in order to face this scenario is not to pursue one of the theories that it includes. A more reasonable choice will consider, on the contrary, the very basis on which the *impasse* is grounded, i.e. the presuppositions and assumptions that originate it. In terms of this

---

[44] It should be read as saying something like *the way in which each meta-ethical theory accommodates these differences).*

general approach, once we have reached a situation in line with (3A), we do not need a new theory but rather, as Smith has recently noted, "(…) a pause while we all take stock"[45]. More generally, once the *impasse* is assumed as a possibility, I believe that four paths are available:

(a) We might simply argue that the *impasse*, although methodologically possible, does not reflect the current stage of development of *Meta-ethics.*

(b) We might simply accept the reality of the *impasse*, and embrace some sort of methodological scepticism about the overall prospects of *Meta-ethics.*

(c) Although real, we could minimize the strength of the *impasse* by simply assuming that one theory will be, in fact, a more plausible or desirable methodological choice *all things considered* – because, let us say, it would avoid some undesirable practical consequences.

(d) Finally, we might well accept the reality of the *impasse* without being committed with any variety of scepticism on the prospects of *Meta-ethics.*

I believe that we should explore the possibilities associated with (d). As result of it, the present work will be concerned with assessing the viability of (d) along a particular route. As I see the point, although current *Meta-ethics* may face a sort of *impasse* (i.e. although two opposite meta-ethical theories may accommodate with an similar degree of accuracy the whole set of platitudes defining our moral stance) *Meta-ethics* should be able to offer some *positive* insights about the nature of our moral thought by reflecting about the proper approach for some fundamental questions about reasons and rationality.

---

[45] Smith, M. 2005. p. 6

4. Looking back and looking forward

In this chapter I have presented an historical overview of the main lines of development of *Meta-ethics*. In my favoured account, *Meta-ethics* has evolved around a complex methodological ideal (*accommodation*), and has supplemented it with the addition of several second-order concerns and a progressive endorsement of the external constraint posited by naturalism. In a certain way, the platitudes that define our moral stance should be explanatory integrated into the set composed by these second-order concerns, keeping an eye on the naturalistic pedigree of our explanations.

After depicting this process of development, I noted that a certain scenario emerges as a possibility associated with the structural image of *Meta-ethics* just referred. In such scenario two meta-ethical theories can accommodate with the same level of accuracy an identical set of platitudes about our moral stance. Leaving aside the effective reality of this scenario, I have suggested that its simple conceivability has explanatory benefits when we try to understand the current shape of *Meta-ethics*. The situation I have located here under the general label of *impasse* has been sometimes referred by some contemporary meta-ethicist as the best way to describe the current stage of development of *Meta-ethics*.

I assumed four possible routes that could help us to face the *impasse*. From chapter 3 to chapter 5 I will be concerned with two attempts to overcome the *impasse* by extracting some meta-ethical conclusions from an analysis of our concept of *reason* as it appears in *explanatory* and *justificatory* contexts. Although both arguments use the same route as me to overcome the impasse - (d) - I will try to establish that they cannot offer a good methodological option to solve the assumed *impasse* in *Meta-ethics*. As far as they want to offer a solution to the *impasse* by offering a direct argument in support of a certain meta-ethical view, these arguments are not going to work. And this is so because they cannot even formulate their favoured arguments without facing a deep level of disagreement (chapters 4 and 6). After presenting a negative standpoint about these arguments, I will introduce the option that I will favour in order to solve the *impasse*, i.e. my favoured interpretation of (d). It moves around the stance of *Meta-rationality*. If I am right, we could extract second-order conclusions about moral judgments by asking second-order questions about our judgments of reasons and rationality (chapters 7-8).

Importantly, I have presupposed a big *as if* in my appeal to a further shift in *Meta-ethics*. All across the board I have implicitly assumed that we can explore the prospects for a meta-normative rooted approach to *Meta-ethics* by simply presupposing the conceptual possibility of an *impasse* - even when the reality of the *impasse* might be decided on more or less empirical grounds. For this reason, my argument for a new meta-normative approach to *Meta-ethics* rests, so to speak, in the conditional plausibility of an *impasse*. Although I will assume this dependence in the current investigation, I tend to believe that even if it were showed that the *impasse* is not real in the end, doing *Meta-ethics* by departing from some meta-normative issues about reasons and rationality could be still the best strategy to obtain some direct answers about the status of our moral opinions.

Chapter 3.

Meta-Ethics and Motivating Reasons

1. An overview of the place of motivational issues in *Meta-ethics*

2. Explaining away the motivational reliability of our moral opinions

3. Toward a qualified version of *motivational internalism*

3.1 Improving *internalism* by distinguishing several possible focal points

3.2 Improving *internalism* by distinguishing the effective source of motivation

3.3 Improving *internalism* by distinguishing the strength of motivation

3.4 Improving *internalism* by distinguishing some motivational shortcuts

4. Looking back and looking forward

Chapter 3

In this and the next chapters I am going to explain why a general meta-ethical strategy (one encapsulated in several *motivational-based arguments*) cannot give us a direct response to the alleged *impasse* surrounding *Meta-ethics*. I will proceed along a quite simple line. First, I am going to ask if the different instances of the argument could be properly formulated. Second, I am going to defend that at least one of the traditional premises included in *any* argumentative instance supporting *the motivational-based strategy* is susceptible to provoke a deep disagreement among those appealing to these arguments – because it is underdetermined or underdescribed. Finally, I will end by assuming that as far as the prospects to end the disagreement around the disputed premise are quite low, *any* motivational argument will not be able to offer, in the end, a direct way-out to the accepted impasse surrounding *Meta-ethics*. While in this chapter I will be focused on the first and second stages of my strategy, in the next chapter I will illustrate why I think that the disagreement around one of the premises compounding the motivational strategy cannot be resolved.

1. An overview of the place of motivational issues in *Meta-ethics*

Meta-ethics has traditionally appealed to motivational issues to defend second-order claims about the status of our moral judgments[46]. Some meta-ethicists have claimed, in fact, that by paying attention to the nature of moral motivation we could secure certain theses about the psychological status of our moral opinions and about their alleged objectivity [Zangwill, N. 2006. p. 262. Wallace, R. J.2006b. p. 191 and Scanlon, T. 2007a p. 7]. In this section I will present a brief overview of the way in which different meta-ethical theories have used certain features associated to moral motivation to secure their grounds.

In my favoured sense, the appeal to motivational issues in recent *Meta-ethics* has moved around three basic steps. First, all motivational oriented approaches to morality have assumed a basic intuition about the type of connection established between our moral opinions and our motives. Secondly, all these approaches have focused on some questions about the nature of

---

[46] In what follows I will understand the term *judgment* along a neutral interpretation. In my intended sense (see Wedgwood, R. 2007. p. 23 and Zangwill, N. 2007. p. 92) by *judging that p* we are neither necessary believing that *p* nor exercising a volitional act with respect to *p* – assessing whether p is the case. *Judgement* stands for whatever mental state we are expressing by sincerely uttering *p*.

our motivating reasons. Finally, all motivational based approaches have tried to use the findings of the theory of motivation to highlight the nature of our moral opinions by appealing to the connection established between moral opinions and motives [Nagel, T. 1970. Chapter 1. Wallace, R. J. 2006/1990]. Dancy refers to this strategy under the term *triangulation* [Dancy, J. 1993. Chapter 1].

The above line of interpretation can be summarized by presenting an argumentative structure or schema usually appealed to ground meta-ethical questions on motivational issues:

> *(1) [A conceptual claim about moral opinions]:* Moral opinions are reliably connected to motivation

> *(2)* [*An substantive claim explaining the nature of human motivation]:* A's motivation can be explained by X-psychological factors or dispositions

> *(3) [A methodological gambit]:* If (1) is conceptually sound, then any second-order approach to the nature of moral opinions should take into account the sort of findings proper of (2).

There are two general points that should be noted about this structure and about how I will use it in what follow.

The *first thing* to note is that the *argumentative force* of the structure rests on the particular content assigned to each one of the components of the structure, as far as it is neutral about any meta-ethical theory. So, even if the structure has no argumentative force as it stands, its soundness in a particular case will depend, for instance, on the particular interpretation that gives content to the general intuition that stresses the motivational reliability of our moral opinions. And the same goes for the general assumptions about the nature of our actions. Depending of how we understand it we could achieve a more or less plausible argument in support of a given meta-ethical view. The overall plausibility of the above structure is conditional to the plausibility associated with our choices to interpret each structural component of the structure.

The second thing to note is that a particular style of rejection of any strategy based on the general structure is open to us as a result of the previous point about the conditional soundness associated with the general structure. In terms of my favoured strategy, we do *not* need to prove that any particular version of the structure above is false in order to reject the

motivational-based argument. Instead, it would be sufficient to reject the motivational-based strategy to prove that *one* of the general assumptions among those included into the structure going from (1) to (3) is *essentially underdescribed*. In my sense, an *underdescribed* assumption could be characterized in the following way:

> **[U]** An assumption *p* located into a general argumentative structure E is *underdescribed* only if we do not have an instance of *p (p')* able to motivate a agreement in support of any particular argumentative structure E' - one where a positive conclusion is entailed from *p'* along with other premises (*q', r', ... n*)

The question to ask now is pretty clear. Which assumption, among those compounding our general structure, is *underdescribed* in my favoured sense? In what follows (see especially chapter 4) I will argue that the assumption that has the property expressed by [U] in the general structure is the first one - (1). To assume that (1) is *underdescribed* is equivalent, in my sense, to reject the possibility of an agreement around any interpretation of it.

But what is (1) saying? (1) has been traditionally understood along two fundamental and contradictory lines of interpretation. *On the one hand,* some people assume that the reliability that (1) defends could be explained by focusing on the conceptual rules governing the use of our moral terms. Moral opinions or moral judgments, they defend, would give rise to motivations in human agents as a matter of conceptual or *internal* necessity.

On *the other hand*, some people reject that the reliable connection between morality and motivation in human agents expressed by (1) could be explained by attending to a certain conceptual feature of our moral opinions. Instead of it, they argue that the reliability proper to our moral opinions is entirely dependent upon external facts related to our motivational dispositions or, alternatively, to those motivational paths that society has enforced to secure cooperation among individuals. In essence, there is a reliable connection between moral opinions and motivations only because a contingent fact (the presence of certain desire, a desire socially rooted) enables our motivational response to those moral judgments endorsed by us.

Nagel summarized both options in the following terms:

> "The names 'internalism' and 'externalism' have been
> used to designate two views of the relation between
> ethics and motivation. Internalism is the view that the
> presence of a motivation for acting morally is guaranteed
> by the truth of ethical propositions themselves. On this
> view the motivation must be so tied to the truth, or
> meaning, of ethical statement that when in a particular
> case someone is (or perhaps merely believes that he is)
> morally required to do something, it follows that he has a
> motivation for doing it. Externalism holds, on the other
> hand, that the necessary motivation is not supplied by
> ethical principles and judgments themselves, and that an
> additional psychological sanction is required to motivate
> our compliance" Nagel, T. 1970. p. 7[47]

But if my general point is plausible - that (1) in the general structure is underdescribed - and, as a matter of fact, *internalism* and *externalism* are exhausting the space of possible interpretations of (1), then any argumentative support offered by both substantive positions will be essentially flawed. If I am on the right track, both substantive theses - *internalism* and *externalism* - should be rejected when they are used to ground *any* particular argumentative instance coming from the general structure – one in support of a wider meta-ethical view. These arguments might be assumed to be not very conclusive at the end of the day.

But, of what arguments are we speaking here? Let me answer this question by offering a standard interpretation of the general structure above depicted. It endorses a non-cognitive conclusion about the status of our moral judgments:

(1) Moral judgments whose content is 'ϕ is right' are necessary connected to a motivation to ϕ-ing (*internalism*)

(2) A's motivation to ϕ-ing is constitutively associated with the presence of a pro-attitude (i.e. a desire) in A to ϕ (*humean theory of motivation*)

---

[47] For classical statements about the motivational divide see Falk, W. D. 1948 and especially Frankena, W. 1958 Both, Wallace, R. J. 2006/1990 and Dancy, J. 1993 note explicitly the importance of Nagel in the current debate. A recent and very influential treatment is found in Smith, M. 1994. ch 4.

Therefore, moral judgments whose content is 'ϕ is right' are expressing, by (1)-(2), a pro-attitude toward ϕ (i.e. a desire, an universal command, a commitment with a system of norms, etc) [Hare, R. M. 1952. 69-73. Railton, P. 2004/1986. p. 12 and 20041986a. p. 47][48]

I will denominate this argument *the classical argument for non-cognitivism* (CA). I am going to comment some variations on CA mainly because they offer a general framework for an overview of the alternatives positions on the nature of moral motivation and moral psychology.

Let me point something brief about the intended *scope* of CA. Although CA-type's arguments could also be adduced in support of a relativistic moral theory – by adding an alternative premise that stresses the variability of human concerns – the argument I have delineated above is aimed to ground a thesis about the type of mental state underlying our moral judgments. In my terms, CA supports basically a psychological conclusion, not a semantic or epistemological one – although we have good reasons to believe that by offering an answer to the latter kind of question we could indirectly provide a basis to respond to the latter set of concerns. Once the historical role of CA and its scope are both fixed, an obvious question is left: Which option remains for those not convinced by CA-type's arguments?

A very usual reaction to CA-type's arguments assumes, as I suggested before, that the interpretation of the motivational reliability expressed by the premise (1) cannot be true. Leaving aside by now whether in fact premise (1) expresses a good interpretation of the sense in which our moral opinions are reliably connected to our motives, let us suppose that we were disposed to reject it. Let us suppose further that we are happy with the *humean* theory of motivation encapsulated in premise (2). If we rejected premise (1), and yet we remained convinced of the plausibility of premise (2), we could formulate the following structure:

(i) Moral judgments whose content is 'ϕ is right' are not necessary connected to a motivation to ϕ-ing in *moral agents* (*externalism about moral motivation*)

---

[48] The reconstruction just offered is proposed by Shafer-Landau, R. 2003.121 in slightly different terms. For some methodological remarks on the importance of this argument see Wallace, R. J. 2006/1990. p.16-20 and 2006b. p. 183. Svavarsdóttir, S. 1999. p.166 and Kalderon, M. E. 2005. p. 8-9.

(ii)   A's motivation to ϕ-ing is constitutively associated with the presence of a pro-attitude in A (*humean theory of motivation*)

Therefore, moral judgments whose content is 'ϕ is right' are *not* necessary expressing - by (i) to (ii) - a certain type of desires.

Therefore, because moral judgments are not conceptually linked to a certain type of desires – by (i) to (ii) - it would be conceptually possible that moral judgments were expressing beliefs [Frankena, W. 1958. Foot, P. 2002/1972. Railton, P. 2004/1986. Brink, D. O. 1989]

Those who argue along the lines above are known as *externalist*. As the conclusion of the argument just referred makes clear, externalists do *not* directly endorse a positive claim about the psychological nature underlying our moral opinions. However, it is customary to presuppose that we might well understand their argument as supporting a more positive claim, which defends that moral opinions might be equated to beliefs with a certain evaluative content that needs to be specified. The point that allow us to assume a more positive conclusion rests on the relation established between premise (i) and (ii). Premise (i) claims that a moral judgment does not directly imply any motive to act. At the same time, premise (ii) defends that any motive, if operative at all, must be associated with a desire or a pro-attitude. Externalism could retain the truth of both claims - (i) and (ii) - by assuming that the state being expressed by our moral utterances is a belief.

Now, let us suppose that the externalist argument is supplemented with a further premise that says:

(iii) A moral agents is one whose motivational set includes a desire to do what is believed to be right [Dreier, J. 2006/2000]

Once (iii) is added, we could explain the motivational reliability associated with moral opinions by appealing to a variety of external facts[49] focused on the desires of certain agents, i. e. those who we consider genuine moral agents. Motivational externalism would simply affirm that the connection reliably established between an agent's moral opinions and her

---

[49] Understanding *external fact* as not referring to any component associated with the very moral judgment

motives is not dependent upon certain psychological features linked to the very utterance of the moral opinion, but rather to certain features associated externally to the agent's desires – or the agent's rationality.

But as I have said before, the classic meta-ethical argument that departs from motivational premises (CA) also rests on an assumption that focuses on the nature of our actions, i. e. it assumes that the *humean theory of action* is true. Nevertheless, when the humean assumption is questioned, an option opens for us to resist the conclusion embraced by CA. In terms of this option, it could be possible to question the merits of the humean theory of motivation encapsulated in (2) while we remain faithful to the merits of internalism - premise (1):

(a) Moral judgments whose content is 'ϕ is right' are necessary connected with a motivation to ϕ-ing (*internalism*)

(b) If rationally formed, A's moral judgment - that ϕ is right - is constitutively associated with the presence in A of an evaluative attitude directed toward some positive features of ϕ (*anti-humean theory of value*)

(c) An evaluative attitude toward some positive features of x is functionally equivalent to an evaluative belief (*cognitivism*)

(d) An evaluative belief could be responsible, in a quasi-causal sense, of the pro-attitude or motive involved in A's ϕ-ing (*anti-humean theory of motivation*)

Therefore, moral judgments whose content is 'ϕ is right' could express, by (a)-(c), an evaluative belief (plus a motivated desire) [Nagel, T. 1970. Dancy, J. 1993. Horgan, T. Timmons, M. 2006b].

Others philosophers attracted by both internalism and the humean theory of motivation, have defended an hybrid version of cognitivism, In terms of this version:

(a') Moral judgments whose content is 'ϕ is right' are necessary connected to a motivation to ϕ-ing (*internalism*)

(b') A's motivation to ϕ -ing is conceptually connected to a *desire* to ϕ -ing (*humean theory of motivation + direction of fit's argument*)

(c') A rational agent is equipped with the rational capacity for attitudinal coherence. This capacity secures that the agent's motives (A's motivation to ϕ-ing) are in line with his evaluative predications ('that ϕ is right') [the basic assumption: rationality as coherence]

(d') No other mental state than a belief could instantiate the functional roles that underlie predications whose form is 'ϕ is right'

Therefore, moral judgments ('ϕ is right') are able to express an evaluative belief because of the enabling role played by our capacity to align our attitudes in a coherent way [Smith, M. 1994. 2004/2000 and 2004/1988]

As the previous reconstruction makes clear, when we look for a motivational argument that support a given meta-ethical view in a conclusive sort of way, it would seem as if we were located in the middle of a number of divergent and equally plausible alternatives and we were asked to pick an option. This feeling, however, is mistaken. The problem is not our inability to choose a knockdown argument in support of a certain meta-ethical view. The problem, rather, is the lack of *any* well-supported framework to offer a knockdown argument departing from motivational premises.

The aim of the current chapter and the following one will be to motivate this suggestion, i.e. to argue that no argument coming from motivation can offer a clear and undisputed clue about the type of state expressed by a moral opinion. And the reason in support of this statement is a simple one: the premise focused on the motivational reliability of our moral opinions cannot be formulated without disagreement. This premise is essentially underdescribed[50]. Consequently, if we chose any route among those that compound the motivational strategy to resolve a meta-ethical issue subject to deep disagreement (for

---

[50] So if we were able to prove that at least one of the basic assumptions is essentially ambiguous, then we would have a certain basis to reject any instance of the general argument - i.e. any possible use of such argument in order to ground particular conclusions about the psychological nature of moral opinions. In this sense, the object of my concern will be the premise about the type of modality associated with the reliable connection between a moral requirement and further motivations.

instance, the one referring to the psychological states expressed by our moral opinions), then the thesis to be defended is that such strategy, which has been unquestioned for a long time, is not going to offer a clear pay-off, leaving the impasse unresolved.

2. Explaining away the motivational reliability of our moral opinions

It has been assumed that we are disposed to embrace a sort of general intuition that connects our moral opinions and our motives. This intuition conform the basis for the first assumption in any motivational argument. It said that:

> (**CT**) Moral opinions are *reliably* connected to motivation.

Let me point one thing about the scope of CT. James Dreier has recently noted that the reliability established between our moral opinions and our subsequent motivations "is an alleged fact about ordinary moral thought, a datum for which every meta-ethical theory must account" [Dreier, J. 2006/2000. p. 547]. David Brink, in a related spirit, writes about the importance of CT for our second-level moral theorizing:

> "We should hesitate to accept any meta-ethical or normative theories according to which moral considerations are considerations to which well informed, reasonable people might always be completely indifferent" Brink, D. O. 1989. p. 37-38

Importantly, Dreier and Brink are defending a central claim related to the meta-methodological importance of CT. As they took it, the conceptual truth expressed by CT cannot be understood either as expressing a positive basis in support of a particular second-order claim about the nature of moral judgments nor as defending a particular normative first-order theory. These exclusions, however, are problematic if we do not say something positive instead about the status of CT. Because, if CT is not directly relevant either with securing the

basis for a particular meta-ethical position nor with supporting any concrete normative moral theory, how should we understand the theoretical importance of CT?

In what follows, I will take the conceptual insight that CT expresses as an *explanatory constraint* on any meta-ethical or first-order moral explanation. As such, CT could work as an external constraint to our theorizing about morality that every meta-ethical or normative theory must take into account. By taking CT at face value – by offering a plausible explanation of why CT is true– any meta-ethical or normative theory would meet a general methodological requirement that permeates the entire realm composed by our normative and evaluative thoughts.

Despite of the minimal status I am presupposing to CT, however, this intuition has been sometimes understood along non-minimal lines. In fact, in recent formulations of motivational issues on *Meta-ethics*, it has been implicitly assumed that there is no important distinction between, on the one hand, widely accepted truism about the reliability associated with moral motivation -CT - and, on the other hand, some interpretations of CT aimed to give philosophical content to the truism itself, so it could be as evidence in support of alternative meta-ethical positions. In a way, it has become a very usual practice to assimilate certain *philosophical theories* about the type of connection established between our moral opinions and our motives to the platitude about our moral motivation expressed by CT. From now on, however, I will assume that *any* meta-ethical theory aimed to explain the nature of moral motivation (along with some other important features of our moral opinions) should accept CT. But it does *not* imply that every meta-ethical theory must offer the same philosophical explanation of CT. CT simply requires that meta-ethical theories explain its reliability in a consistent sort of way [see Shafer-Landau, R. 2003. p. 156-157 and Wallace, R. J. 2006b. p. 183 for this general point].

Having distinguished CT from the *philosophical* constructions around it, in the next section I will highlight the content of what we are assuming when we defend, in a philosophical sense, that our moral opinions are *internally* connected to our motives. Once this is established - once it has been delineated what I consider a reasonable version of *internalism* - I will pay attention, in the next chapter, to the main problems lurking around this feasible version of *internalism*. All along this critical process I will delineate what an externalist theory could say about moral motivation. As I announced before, neither internalism nor externalism will be pervasive at the very end.

3. Toward a qualified version of motivational internalism

Motivational internalism defends an apparently simple second-order thesis about the nature of our moral motivation. It says that there is an internal or necessary connection between moral judgments and motivation for action. In appealing to *necessity*, internalists normally appeals to a *conceptual* sense of necessity, one supported by the regularities and expectative that govern our uses of basic moral terms [Wallace, R. J. 2006b. p. 182. Svavarsdóttir, S. 2006. p. 164. Wedgwood, R. 2007. p. 23].

Motivational internalism could be presented by means of a conditional formula:

$$\text{(I) } N \, (A_{MR} \, \phi \rightarrow A_M \, \phi)$$

The formula above is normally read as affirming that, necessary, if A is morally required to ϕ, then A will be motivated to ϕ-ing. Despite of its apparent directness, however, internalism could be narrowed in several ways to clarify better what it is being really stated by (I). For instance, (I) could be narrowed along its *proper focus*, i. e. along the type of items (requirements, moral judgments, etc…) to which (I) is referring. (I) could also be improved by clarifying the type of *mechanisms* that surround the motivational reliability of our moral judgments (whether they are elicited by the very utterance of a moral judgment or whether they are non-constitutively dependent on an utterance of a moral opinion. Finally, the above principle could be modulated by specifying the *strength* of those motives referred by it - whether they are always overriding non-moral motives or whether, in turn, they can be involved in motivational conflicts. Having acknowledged these dimensions, in the next section I will narrow the focus of (I) along them.

3.1 Different formulations along different focal points

It has been widely acknowledged that (I) could express at least two different types of statements. We could subscribe either a *requirement-based version of* (I) or a *judgment-based version of* (I) [Brink, D. O. 1989. p. 40-41. Shafer-Landau, R. 2003. p. 143]. Let me introduce them in order.

Requirement internalism (RI) defends, in essence, that if an agent (A) is morally required to do ϕ then A will be motivated to do ϕ.

(RI) N (If ϕ is morally required from A, then A will be motivated to ϕ)

RI is necessary connecting moral obligations (understood as external facts entirely independent of the agent's attitudes) to the agent's motives. Evidently, there is a basic problem for RI[51], which is related directly with the choice of the items that express the conceptual truth with which (I) is committed. The problem is that it could be possible, as a direct consequence of the modal status of (I), that a moral obligation is *not* applicable to an agent by the simple fact that he is not motivated to act in the way commanded by the moral obligation, To put in other way: When we understand the modality expressed by (I) by means of RI - relating moral requirements as such to an agent's motives - we end by affirming that there is a necessary connection between the *existence* of a requirement and our motives. And this is completely counterintuitive. It is clear to us that some requirements are applicable even when we are not motivated to act in line of what is being demanded by the requirement [Foot, P. 2002/1972. Brink, D. 1989. p. 41. Shafer-Landau, R. 2003. p. 143. Zangwill, N. 2007. p. 94]

As far as any version of (I) would surely share the sense of *obligation* just noted (where the normative status of the obligation is entirely independent of recognition by the agent), (I) must explain the alleged connection between moral opinions and motives in a different way than (RI) does. (I) could be plausibly understood not as affirming that the *fact* that a certain moral requirement applies to an agent is internally connected to his motives, but rather as defending that if an agent *judges* that a certain action is morally right, he will be motivated to act in line with his self-acknowledged judgment [Dancy, J. 2006. p. 133. Shafer-Landau, R. 2003. p. 143]. *Appraisal* or *judgment internalism* (JI) defends this route:

(JI) **N (If A *judged* that ϕ is morally required then A will be motivated to ϕ)**

---

[51] **(RI)** is sometimes denominated *existence internalism* (Shafer-Landau, R. 2003. p. 143) or *agent internalism* (Brink, D. 1989. p. 40)

JI needs a minimal qualification. An agent could judge that doing ϕ is morally required but not to be motivated to ϕ-ing because he does not acknowledge that the requirement was commanding an action *from him*, i.e. he does not believe that he *himself* is morally required to act in a certain way. So, the indexical character by means of which our normative judgments influence our will requires an improved version of (JI)

(**JI** _{indexical}) N (If A judges that he *himself* is morally required to ϕ, then he will be motivated to ϕ)

Leaving aside this proviso [see Burge, T. 1998 for an extended treatment], we surely can agree on the fact that JI, at least at a first sight, is a good candidate to encapsulate the truth contained in (I). Ignoring by now some questions related to the force of the motivation JI is referring to, the basic problem surrounding JI is surely its vagueness. Even if we assume that JI is a plausible rendition of (I), how should we understand JI to better secure the modality assumed by (I)?

On the one hand, JI could be formulated along ***pure psychological terms***. JI would defend that if an agent has judged that he is under a certain obligation then his motive to comply with his self-assumed obligation is caused, so to speak, by the very psychological state that is being expressed by his self-assumed moral opinion. JI would defend, for instance, that the necessary connection established between an agent's self-referred moral opinions and his motives is secured by the causal role of certain psychological states that are underlying the agent's moral utterances. As I will defend later, most formulations of internalism are committed to a psychological reading of JI.

But although JI is able to minimize its essential vagueness by assuming a *psychological* reading, a problem for JI would remain yet. In terms of this problem, the truth of JI would undermine an intuition focused on our motivational powers. In terms of a widespread intuition, moral motivation must be considered, if not as a capacity entirely independent of our psychological capacities, at least as a capacity that could sometimes transcend our *current* psychological capacities. But if JI is true, the motivational import of our moral opinions depends essentially upon certain *current* psychological factors, i.e. those underlying A's judgment. Therefore, JI cannot be true, at least along an unqualified psychological reading [Brink, D. O. 1989. p. 45]

However, there is *second alternative* to reduce the vagueness of JI. In these terms, the connection to which JI is giving voice could be stated *indirectly* by referring to the reasons expressed by our moral judgments. Along this line, JI is assuming that if an agent believes that he is morally required to ϕ, then *there is a normative reason* favouring his ϕ-ing:

> **(JI rationalism)** N (If A believes that ϕ is morally required, then A has a justifying reason to ϕ)

Evidently, those who accept this claim are disposed to support JI rationalism by subscribing a different albeit related thesis, which affirms the existence of an intimate connection between normative reasons and motivation [Wallace, R. J. 2006b. p. 185]. Thus, once we accept JI rationalism, the internal connection favoured by JI will be secured by accepting an additional thesis:

> **(NI appraisal)** Necessary (If A believes that he himself has a reason to ϕ then A will be motivated to ϕ *as far as A is rational*)

(NI appraisal) secures the necessary connection between a self-assumed moral opinion and a motive - (JI) appraisal – by explaining it in terms of our general responsiveness to the features that favour an action, i.e. by stressing our general responsiveness to *reasons* [see Smith, M. 1994. p. 61 for a different account of (NI appraisal) in terms of *rationality* as *coherence*]. Accordingly, JI can be understood as defending

> **(JIappraisal+rationalism)** N (If A judges that he *himself* is morally required to ϕ, then he will be motivated to ϕ *as far as he is rational*)

In terms of the above formula, the internal connection established between an agent's self-referred moral opinion and his motives is mediated by our unqualified responsiveness to the beliefs about our own reasons.

Until here I have presented two options to understand the focus of (I). Now, let me move toward the different question of the focus or effective mechanism that underlies (I).

3.2 Improving internalism by distinguishing the effective source of motivation

I have encapsulated *internalism* into a simple formula:

(I) N (A $_{MR}$ ϕ $\rightarrow$ A $_M$ ϕ)

In the previous section, I presented some clarifications on (I) to state what type of instances support the modality that (I) expresses[52]. I have assumed that JI is a plausible view if properly interpreted. But we should make a set of further clarifications to offer a version of JI maximally plausible. We should ask, for instance, by means of what type of mechanism the connection defended by JI could be established. Or, to put in another terms, we could ask about the type of explanation that should be favoured to give sense to our moral motivation. Along this lines, some philosophers [see Svavarsdóttir, S. 2006. 164. Mele, A. R. 2003. Ch. 4] have defended that a *pure version* of JI would get the conceptual truism expressed by JI in an accurate way. According to *pure internalism*

> **(JI - Pure version)** N (If A judges that he himself is morally required to ϕ, then (i) A will be motivated to ϕ and (ii) the fact that A is motivated to ϕ can be explained by the very utterance of a moral judgment[53]

Non-pure internalism, in turn, will defend that

> **(JI- Non-pure version)** N (If A judges that he himself is morally required to ϕ, then (i) A will be motivated to ϕ and (ii) the motivation to x is not necessary linked to A's evaluative utterance

The main difference between *pure* and *non-pure* versions of JI resides in the role that they propose that our moral opinion plays to cause our motives. Both versions accept the general thesis encapsulated in JI and both offer, at the same time, a different explanation about the internal connection between a moral opinion and an agent's motivation. On the one hand,

---

[52] For the present purposes I will assume that an appraisal version of (I) can be defended without going into details on whether it needs to be supplemented by appealing to reasons.

[53] Because the agent is expressing, by means of his utterance, a desire or an evaluative belief whose content is referring to ϕ

pure versions of JI claim that the psychological state that a moral opinion expresses *effectively conforms the very state of motivation* underlying an action in accordance with the agent's moral opinion. As Nick Zangwill has recently written, "that motivation is *essential* to moral beliefs" [Zangwill, N. 2007. p. 94]. *Non-pure* versions of JI defend, on the other hand, that an agent's being motivated by a moral opinion is secured by the further intervention of a psychological state *other than the state that is expressed by the moral utterance*. So, in the terms favoured by non-pure versions, the state expressed by a moral opinion does *not* directly secure the agent's motives to act in a certain way although the motivational connection is constructed in terms of necessity [Smith, M. 1994. p. 61]. I will assume here that (I) is usually understood as supporting a *pure* version to explain the connection between moral opinions and motives [Svavarsdóttir. S. 2006. p. 164]. As far as I believe that internalism can be understood along with a defence of *rationalism,* I will accept hereafter that the reason-based version of JI is better understood as a *non-pure* version of internalism[54].

3.3 Improving internalism by distinguishing different degrees of motivational strength

Until now it has been argued that, as a matter of conceptual necessity, any moral opinion, if sincerely endorsed, motivates on its own – either in virtue of how the psychological state that it expresses is necessary connected to our motives or in virtue of a general capacity to *respond* to reasons (Wallace) or to be coherent (Smith). This formulation has emerged along successive improvements of (I). In this brief section I would supplement further JI by asking for the motivational *prevalence* of our moral opinions to determine what to do. I will be concerned, in essence, with a single question: What is the degree of *prevalence* or *strength* by means of which the motives internally associated with agent's moral utterances determine our intentions?

---

[54] In Smith's favoured reading, JI is saying that

> JI $_{\text{SMITH}}$ N (If A judges that he himself is morally required to ϕ, then A will be motivated to ϕ *as far as A is rational*)

As far as Smith understands *rationality* as a general disposition to align our attitudes in a coherent way and such disposition is not essentially connected with the utterance of a moral judgment, we must assume that JI $_{\text{SMITH}}$ is a *non-pure* reading of JI.

In terms of this basic concern we could distinguish two different options. It could be defended, first, that moral motives determine our intentional action in a *conclusive* sort of way. In terms of this intuition, if an agent judges that some act is right then he will necessary act in accordance with the motivation underlying his moral judgment [Hare, R. 1981. p. 60-61]. I will refer to this option as **strong internalism** (JI $_{strong}$). JI $_{strong}$ assumes that the following formulation of JI could be plausible:

> (**JI $_{strong}$**) Necessary (If A judges that ϕ is right then A will be *conclusively* motivated to act in accordance with his judgment).

It is assumed that JI $_{strong}$ is problematic because of a simple fact [Watson, G. 2004/1975. 2004/1977. McNaughton, D. 1987. Chapter 8. Zangwill, N. 2007.p. 95]: JI $_{strong}$ does not consider the central role that some motivational disturbances - *akrasia*, *accidie*, or simple *amorality* - play to conform our intentions. It has been widely observed, for instance, that agents could fully endorse a moral opinion (in the precise sense delineated by JI) without being moved to form an intention to act in accordance with their acknowledged moral commitment. But if this is a common fact of our experience, i.e. if it is not only conceptually conceivable but strikingly real that we are sometimes unmoved by our moral opinions, then we should reject either the general framework offered by JI or, alternatively, the strong reading of this framework picked out by JI $_{strong}$.

In my opinion, as far as that both sort of facts - JI and the commonsensical phenomenon of *akrasia* - conform to solid intuitions, we should look for a *weaker* formulation of JI, which includes the central role played by those motivational disturbances in the account of the motivational import of our moral opinions. A plausible route to include the role played by these shortcomings could stress, as it has been recently argued, the *conditional* prevalence by means of which our moral motives determine our intentions. In this route, although a moral opinion would always express a motive (M) to be accounted for when forming an intention to act, the *overall* prevalence of M is always dependent on the absence of stronger motives – M_1, $M_2, ...M_N$ [Brink, D. O. 1989. p. 41. Shafer-Landau, R. 2003. p. 142]. In essence:

> (**JI $_{weak}$**) Necessary (If A judges that ϕ is right, then A will have *some* motivation to act in accordance to his judgment)

Evidently, by defending JI $_{weak}$ the previous objection could be easily debunked. Normal cases of *akrasia* could be understood now as cases where the agent's moral motives are playing a motivational role even when they are not *unconditionally* prevalent in determining the agent's intention. In sum, we could argue that the utterance of a moral opinion in such cases is also reliably followed by a motivation to form an intention to act although such motive is not necessary prevalent in an overall sense. So, even if JI were true (and moral motivation can be internally constructed) it does not imply that our moral motives are guiding our choices or intentions unconditionally [Nagel, T. 1971. Wallace, R. J. 2006b. p. 188][55]

3.4. Improving internalism by distinguishing some motivational shortcuts

Once we consider the degree of motivational prevalence associated with our moral judgments, which would be the best formulation of JI?

Take the following one:

A qualified version of (I):

[IQ] Necessary (If A judges that he himself is morally required to X then (i) he will be motivated, to some extent, to X and (ii) his being motivated to X could be explained either by appealing to the nature of A's evaluative judgment or by appealing to A's *rationality*)

IQ is committed with a wider claim, which assumes that, necessarily, an agent is disposed to form a motive to act in line of his self-referred moral opinion[56] even though such motive will not always be prevalent in determining his behaviour in an overall sense [Blackburn, S. 1998. p. 61]. So, despite of the fact that A's opinion that φ-ing is morally right should always be a relevant feature to be accounted for in explaining A's actions, the most plausible reading of this explanatory proviso should assume that A's moral opinions are playing a motivational role *only in a conditional way*.

---

[55] The reason-based version of JI (JI$_{appraisal + rationalism}$) could also explain those cases where A is not motivated to φ even if he believes that φ is right. Accordingly to this version, A is simply behaving irrationally in such situations. A is not weighing the considerations that would justify his behaviour appropriately. See Scanlon, T. 1998. p. 20-22. Shafer-Landau, R. 2004. p. 144. Wallace, R. J. 2006b. p. 187. Wedgwood, R. 2007. p. 25.

[56] In my sense, 'in line with' implicitly refers here to the way in which the strength of A's motives should be proportional to the degree of confidence of A's judgment. For two detailed accounts of the importance of this qualification in motivational debates see Smith, M. 2004/2002 and Zangwill, N. 2007, especially at p. 95-96

We could surely supplement IQ in a further direction by *explicitly* taking into account the influence played by some negative psychological conditions in our understanding of IQ. Remember that in the previous section I modified JI to accommodate the important role played by certain motivational disorders. I assumed that we must understand the motivational importance of our self-assumed moral opinions in a conditional sort of way – at least if we want explain certain cases of moral motivation. But the thing is that it would be conceptually possible to cancel the connexion established between our moral opinions and our motives *even before* any deliberative process takes place. It would be possible, in sum, that in virtue of certain motivational deficiencies the agent was not even able to be motivationally responsive (at any degree) in light of his self-assumed moral opinions [Svavarsdóttir, S. 1999. p. 164. 2006. p. 164]*.* But if such cases are accepted as exemplifying possible scenarios, any sense of prevalence attached to moral motivations for human actions should be bracketed, leaving some space for an alternative formulation of the motivational import of our moral opinions. An even more qualified version of IQ could be formulated then by simply considering such cases:

> **An even more qualified version of [IQ]***:*
> Necessary (If A judges that he himself is morally required to X then (i) he will be motivated at some extent to X and (ii) A's being motivated to act in such way can be explained either by merely appealing to A's normative judgment or by referring to A's rationality - *except in individuals exemplifying an abnormal motivational make-up.*

I will assume hereafter IQ *modified* as the most reasonable way to unfold the content of IQ. IQ is my favoured version of (I).

4. Looking back and looking forward

Now let me recap. IQ *modified* has been extracted from a general methodological intuition. The intuition assumes that to define the precise content of (I) we should consider several dimensions along with which we usually speak of moral motivation. We should choose the most reasonable version of (I), in virtue of how it would take into account these dimensions.

The resulting version will offer an initial best statement of the motivational intuition that appears as first premise in any motivational-based argument aimed to establish a meta-ethical conclusion about the psychological nature of our moral judgments. So, IQ *modified* – IQ hereafter - has been formulated by taking four different dimensions into account:

- (i) What is the proper *focus* that grounds the modal claim defended implicitly by (I)?

- (ii) How such conditional schema is supposed to be secured? Are we assuming an *immediate or pure* sort of connection between moral opinions and motivations or are we speaking of a *mediated* or *non-pure* link here?

- (iii) What is the strength of the motivational state in the overall context of practical reasoning?

- (iv) In which type of *deliberative* framework should we include IQ?

As I have just said, IQ offers the most reasonable response to these sorts of questions. In the next chapter I will review some problems that any motivational argument faces – even when it is resting on IQ.

Chapter 4

Exemplifying the *impasse*: moral motivation

1. Conceptual arguments against IQ

1.1 Brink's version

1.2. Mele's version

1.3 Shafer-Landau's version

2. Explanatory arguments against IQ

3. Securing internalism?

3.1 Improving the *conceptual* strategy

3.1.1.   Hare's version

3.1.2.   Problems for Hare's version

3.2 An *explanatory* rejoinder

3.2.1.   Smith's *moral fetishism argument*

3.2.2.   Problems for Smith's *moral fetishism argument*

4. Looking back and looking forward

Chapter 4

So far I have argued that once it is assumed that *Meta-ethics* is facing a sort of *impasse* (chapter 2) we could propose an argumentative strategy to overcome such *impasse*. We could propose something close to the trio of assumptions below to minimize the meta-ethical *impasse* (chapter 3):

(1) *[A conceptual claim about the motivational import of our moral opinions]:* Moral opinions are reliably connected with motivation

(2) *[An accepted theory explaining human motivation]:* Non-moral motivations can be explained by X-psychological factors or dispositions.

(3) *[A methodological gambit]:* If (1) is conceptually sound, then any second-order approach to the nature of moral opinions should take into account the sort of findings proper of (2).

Obviously, it could be argued that the whole argumentative enterprise that (1)-(3) represent is ill conceived from the very beginning because of the extreme complexity underlying our moral thought. Such complexity (the extreme variety of psychological items to be included in any realistic reconstruction of our moral stance) could cancel any approach based narrowly on a single argumentative schema [see Railton, P. 2006a. p. 15 and 31 for a related point focused on the general phenomenon of *normative guidance*].

But I will argue hereafter that we should not be impressed by this objection. In turn, we should be aware that even if we were accepting this psychological complexity as a feature defining our moral thought, it would still be useful to know which type of psychological category (cognitive or non-cognitive) plays the main functional role in structuring our moral utterances. On the one hand, by knowing the type of psychological state that we are expressing in a moral sentence we could gain, as I assumed before (chapter 2), an additional route to answer some classical questions about the status of our moral opinions - ontological and epistemological ones. On the other hand, by isolating a psychological category we could facilitate a conceptual framework to narrow any empirically oriented approach. Leaving aside these questions, let me return to the structure depicted by (1)-(3). In chapter 3 I advanced a

problem in relation to this structure, which was prior or independent to any particular argument derived from the structure. The problem concerns the status of the argumentative structure and it basically notes that, even if we were accepting its explanatory force to derive particular responses about the psychological status of our moral opinions, we should agree on the fact that *at least one of the general assumptions included in (1) to (3)* expresses an essentially *underdetermined* or *underdescribed* claim.

By *underdescribed* I understood a property quite specific, which was attached to the content expressed by a general assumption in line with (1) or (2). I suggested that an assumption (A) is *underdescribed* only if it cannot be *formulated* under any particular and undisputed interpretation, i.e. one aimed to unfold the precise content of what is being assumed by A. So, if it was proved that at least one of the assumptions that conform the structure going from (1)-(3) is *underdescribed,* we could obtain an argumentative tool against those who use this general structure as a meta-ethical argument in support of their views. Or to put in another way: As far this general structure is damaged in virtue of the dubious status of one of its core assumptions, the plausibility of any particular argument resting on it should be re-evaluated and, if I am correct, severely questioned.

But how should we argue to support of the *underdescribed* character of a given assumption (A)? To answer this question I suggested a simple route. I defended that if

> **(a)** *The most plausible interpretation* of A were the focus of a deep disagreement (one exemplified by means of a set of critics focused on A)

and

> **(b)** A better alternative to A does not emerge from such disagreement

then we should end by assuming that A, if not dubious by itself, is surely unable to support any particular knockdown argument derived from the general structure.

As I announced in chapter 3, I believe that (1) in the general structure above is the assumption to question Evidently, in the terms that (a) and (b) favour, to question (1) is equivalent to introduce, firstly, a maximally coherent version of (1) and then asks, at a second stage, for the prospects of such versions. So, if in the previous chapter I introduced what I

took to be, if not the most reasonable, at least the most extended account of (1) - IQ - in the current chapter I will be concerned with a set of critical assessments of IQ, which suggest a deep disagreement around (1). The aim of this chapter will be to show how (1) expresses an essentially underdescribed or underdetermined assumption as a direct result of a dialectical process of successive criticism and rebuttal. This assumption is unable to support any motivational argument in a knockdown-style, i.e. as a direct route that supports a meta-ethical conclusion about the type of psychological state that our moral utterances express.

1. Conceptual arguments against IQ

Let me remind you the interpretation of (I) supplemented in the previous chapter that I labelled as [IQ]. There I assumed that

> **[IQ]** Necessary (If A judges that he himself is morally required to X then (i) he will be motivated at some extent to X (ii) A's being motivated to act in such way can be explained by appealing to A's normative judgment - *except in individuals exemplifying an amoral make-up*

IQ has been formulated after successive improvements on (I), being formally equivalent to (I). We should accept that because of the fact that (I) is committed with

$$\text{(I)} \quad \text{Nec} (A_{MR}\ \phi \rightarrow A_M\ \phi)$$

and

$$\text{(I)} \quad \neg\text{Pos} (A_{MR}\ \phi \wedge \neg\ A_M\ \phi)$$

IQ can be falsified if A is judging that he himself is morally required to X without being motivated at some extent to act in line with his moral opinion - in absence, again, of any serious motivational failure. But once we have noted this, it should be easy to see how those who reject IQ are proceeding at this stage. They are taking advantage from the formal

equivalence between (I) and IQ to propose the following argument [Svavarsdóttir, S. 1999. p. 176 and 2006. p. 165. Mele, A.2003. p. 113):

(i)   If internalism is understood as meaning something close to IQ, then it should not be possible, neither in a conceptual nor in a psychological sense, to be unmotivated in the presence of an acknowledged normative belief – leaving aside agents with motivational disorders.

(ii)   It is conceptually possible to conceive cases where the agents remain unmotivated in presence of a normative belief and, besides, there are situations where the agents do not take into account moral considerations to form heir intentions.

Therefore, (IQ) must be false because it is neither conceptually nor psychologically necessary to be motivated at some extent in the presence of a first-oriented normative belief

Let me say something the argument sketched above.

Premise (i) simply establishes what it is required to reject *internalism*. As Shafer-Landau has recently pointed " (…) establishing the *possibility* is all we need to undermine motivational judgment internalism" [Shafer-Landau, R. 2003. p. 145. Although see Blackburn, S. 1998. p. 61 for a less optimistic view]. Premise (i) is assuming that any interpretation of the motivational import other than IQ will involve a *conceptual* mistake, so that it would not be conceptually possible to offer an interpretation of this motivational link other than those delineated by IQ - at least if we are respecting the practices ruling the use of our basic normative terms.

Externalist arguments are mainly focused on establishing the falsity of internalism by any of the possibilities pointed out by (i). Consequently, the arguments supporting externalism could be roughly separated in terms of the type of possibility that they purport to actualize. The aim of some classical arguments for externalism is to expose the reality of certain psychological characters, pressing on the intuition that such psychological cases, if correctly depicted as real, are defying IQ. In other cases, by contrary, the aim of the argument will be to show that, once the set of conceptual features posited by IQ are unfolded to explain our moral motivation, some deeply entrenched intuitions would emerge as conceptually incoherent.

From now on, I will explain with some detail how externalism secures the most important premise in its case against IQ - premise (ii). I will first expose Brink's argument for the *conceptual* possibility of cases in which an agent is fully endorsing a moral claim without being motivated to act in accordance with it. Afterwards I will present a more promising option, one defended by Russ Shafer-Landau and Alfred Mele. Then I will focus on a methodological-based argument in support of externalism posited by Sigrún Svavarsdóttir.

1.1 Brink's version

David Brink offers an argument aimed to reject *internalism* by confronting it with a familiar character that inhabits our commonsensical views about virtue and moral worth [see Arpaly, N. 2004. Chapter 1 for a more nuanced view]. In Brink's sense, *internalism* - understood along the lines favoured by IQ - implies the conceptual impossibility of *the amoralist*, i. e. of someone who recognizes that a moral consideration applies to him but remains completely unmoved by his own recognition[57]. As far as it is possible to *conceive* an amoral person, Brink argues, *internalism* should not be accepted as an accurate second-order thesis about moral motivation [Brink, D. O. 1989. p. 46. Compare with McNaughton, D. 1988. p. 135 and Smith, M. 1994. p. 66][58].

Although Brink's argument is direct, its fundamental problem is that it does *not* offer additional reasons in support of the central assumptions that ground its overall case against internalism. In Brink's sense, the *conceivability* of a motivational type (amoral person) falsifies the conditional endorsed by IQ. But the thing to ask, surely, is whether this general appeal to *conceivability* is itself out of the scope of further criticism [Brink, D. O. 1989. p. 48]. Because, if Brink is right, it would seem as if by showing the *conceivability* of certain scenarios where a given concept (M = moral endorsement) is apparently used to describe what

---

[57] Without being the case that an intervening factor is cancelling his motivational strength – as IQ modified is noting

[58] IQ is being understood in disjunctive terms. Accordingly, the fact that there is a necessary connection between agent's moral opinions and agent's motives can be explained either in virtue of the type of psychological state underlying agent's moral utterance or in virtue of agent's rationality. If so, two types of counterexamples could be posited against IQ. First, it could be suggested the possibility of the amoralist. Secondly, and here goes the point of this footnote, it could be suggested the possibility of an agent who being able to acknowledge an undefeated reason to φ, is completely unmoved to φ. While some internalists - those assuming NI $_{appraisal}$ - should be committed to deny the possibility of this character, some externalists forcefully stress our common ability to be unmoved by reasons in order to deny IQ. Thus, although the current section is devoted to the *amoralist*, it should be assumed hereafter that we could also reject IQ by stressing the just referred phenomenon. See Foot, P. 2002/1972 for the classical source and Greenspan, P. 2006. p. 181-182 for a more recent discussion.

is going on (*A believes that he himself is morally required to* ϕ *but A is not motivated to* ϕ), we could derive a conceptual feature guiding the use of M in any context[59]. But surely, before extracting consequences about our uses, we should question how to interpret the conceptual insights that arise from the *conceivable* scenarios where M is appealed to determine the actual *content* of M.

So, let us suppose that it were possible to *conceive* a situation in which an agent endorses a moral opinion while he remains unmoved by his self-assumed moral commitment. If so, Brink assumes that it would be feasible to locate, among the components that define the *actual* content of M, a feature that allows us to use M in situations where an agent recognizes a moral demand on his own actions while he remains unmoved by his own self-assumed moral endorsement. As I see the point, the thing to ask at this stage is if we are really entitled to rest on intuitions about cases where we could *conceivably* apply C to derive conclusions about the *actual* content of C. Surely we are not entitled to offer a proper answer to this concern unless we take a clear stand in a wider debate that focuses on the relationships established between *conceivability* and *conceptual possibility*. But if this is so, then a genuine debate about the motivational force of our moral opinions is reduced to a metaphysical discussion, i.e. to a discussion about the extent to which the content of C is determined by some conceivable scenarios where C is used. Surely it should be a bad thing for motivational arguments to be reduced to metaphysical arguments, which move around the relation established between *possibility* and *conceivability*.

Thus, *internalists* do not claim that the *amoral* agent is not conceivable. They could accept that such motivational exemplar is conceivable without assuming that he could offer some kind of insight about the actual content of our central moral concepts. So, while Brink's argument defends that internalism is false because it makes *conceptually* impossible something easily *conceivable* (*amoralist*), those committed to internalism could suggest that Brink's central intuition needs further justification. It needs basically to secure the link between *conceivability* and *genuine conceptual* possibility. The thing to ask now is clear: Could we have an alternative conceptual argument against IQ that does not rest on such strong metaphysical assumptions?

---

[59] Shafer-Landau encapsulates Brink's strategy in the following way: "(…) What is imaginable is conceptually possible. The amoralist is imaginable. Therefore the amoralist is conceptually possible". He focuses later on the point I am trying to put on the foreground: "(…) We surely can imagine a person uttering the words 'that action is morally right', while remain unmoved. The question is whether we can imagine a person sincerely saying such a thing". Shafer-Landau, R. 2003. p. 146

1.2 Mele's version

An option to supplement Brink's conceptual rejection of IQ that I want consider here has been recently developed by Alfred Mele [Mele, A. R. 2003]. Mele starts by considering cases of *moral listlessness*. In such cases, an agent is apparently able to judge that certain act would be morally required even if she is unmotivated to act in accordance with such moral opinion. An important feature of Mele's examples is related to the character of the agent-type involved in his favoured scenarios. In Mele's proposal we do not find an *amoral* person. On the contrary, the agent-type who fails to be moved by his self-assumed moral judgment has proved to have a consistent disposition to behave morally in the past. Thus, Mele assumes that to evaluate the above noted cases it would be better if we simply rejected from the very beginning the assumption that affirms that *amorality* or *wickedness*, understood as an agent's character-traits, would explain his observable behaviour. As far as we know that this person is capable of genuine moral motivation in other occasions, this interpretation must be ruled out from the start [Mele, A. R. 2003. p. 109-111 and 117].

Mele notes that in some cases of depression an agent is able to fully endorse a moral opinion while remaining unmotivated to act in the way that his assumed moral conclusion requires. Mele focuses on Eve:

> **[Eve's case]** "Consider an unfortunate person – someone who is neither amoral nor wicked – who is suffering from clinical depression because of the recent tragic deaths of her husband and children in a plane crash. Seemingly, we can imagine that she retains some of her beliefs that he is morally required to do certain things (…) while being utterly devoid of motivation to act accordingly (…) If we can imagine this (more precisely, if it is possible) then belief internalism and strong internalism are false: agents may believe that they are morally required to A and yet have no motivation to A" Mele, A. R. 2003. p. 11

Mele is concerned, again, with a variety of conceptual question:

> " I concentrate on the specific question whether is conceptually and metaphysically possible for Eve to

> continue to believe that she is morally required to aid her
> uncle while being utterly unmotivated to aid him (…)"
> Mele, A. R. 2003. p. 113

And he is going to argue that such cases are conceptually possible only if we locate them into a wider image focused on the nature of our evaluative beliefs. In Mele's terms, as far as we all endorse a common image about how our *evaluative* beliefs are accepted (a folk-theory in line with a more general view about how *any* belief is accepted), we could be in a position to offer a direct response to the previous question supporting the conceptual tenability of the scenarios favoured by externalism. And we could do such a thing without assuming a heavy-weight metapysical distinction, as Brink did before.

Mele starts by assuming that evaluative beliefs can be characterized as having two basic features shared by non-evaluative beliefs. *The first feature* assumes that evaluative beliefs are susceptible to be depicted as having *representational content*, i.e. one able to be grasped in propositional terms [Velleman, J. D. 2000b. Sinclair, N. 2007].

The second feature is somehow related to the previous point. It assumes that the psychological state that we exemplify when we *endorse* or *accept* a certain content (p) along with a belief-mode of *acceptance*, should be directly responsive to the truth or falsity of *p*. So, believing that *p* gives voice to a psychological state guided by a certain aim, which requires that you, as believer, pay attention to the set of truth-making features associated with the propositional content of your beliefs. Evidently, we could attain this end either by certain sub-personal mechanisms or by referring to a personal *policy* aimed to obtain evidence, in any situation, about the object of our inquiry. But the important thing is that beliefs aim at truth in one way or another [Williams, B. 1971. Bratman, M. 2001].

As I noted before, Mele assumes that our evaluative beliefs are governed by roughly the *same* rules of endorsement associated to non-evaluative beliefs, i.e. by rules oriented to focus the agent's deliberation and other cognitive capacities such as memory and attention on the aim of gathering evidence that would make more probable the truth of a certain *content* – an evaluative content. But once we accept this general framework, how is it supposed to support a conceptual attack on IQ?

Let me remind you that we assumed before that externalists are searching for a plausible story to deny the central claim defended by IQ (that it would *not* be *possible* to sincerely affirm that oneself is morally required to do a certain thing and, at the same time, to remain

unmoved to act in line of our judgment). Accordingly with it, Mele offers a plausible rejection of the central claim defended by IQ - which rests on a minimal thesis about our beliefs - without appealing to whether certain scenarios are conceivable or not. Let us take Eve's example again, a case where a woman in grief by the death of her husband and children claims to be under the moral obligation to help her uncle but, at the same time, is clearly unmoved by such belief. Eve is now in a medical examination three or four months after the accident. When she is asked whether she is morally required to help to her uncle she responds, after evaluating certain evidence that she is. So, it seems as if Eve, when asked about her belief in a moral obligation in respect to her uncle's care, were examining certain evidence – about her uncle's conditions, her availability in terms of time, and so on – and only after such process has finished she is disposed to endorse a certain content if asked about what she is believing. In this respect, Mele argues, our moral beliefs share a basic feature with our non-moral beliefs [see Timmons, M. Horgan, T. 2006b for further commonalities]

Mele assumes that if Eve's evaluative belief is understood along a line of continuity in respect to other, non-evaluative, beliefs, then it could be possible to explain why Eve is endorsing such evaluative belief while she remains unmoved to act in accordance with her evaluative belief. In terms of this minimal story about what a belief is - one partially encapsulated by the previous features – we should separate, in assessing a belief, those questions concerned with the rules governing the endorsement of a certain propositional content from those processes related with agent's motivation. Even if we could facilitate a conceptual argument that favours the necessity of certain rules of endorsement (truth, for instance), whether an agent who believes certain content is motivated to act in accordance with the believed content is a contingent fact, which is related to the agent's motivational capacities [Mele, A. R. 2003. p. 114].

But if the above point is on the right track [Smith, M. 1987. Little, M. O. 1997], Eve could endorse an evaluative belief - taking as ground certain amount of evidence, as we do in non-evaluative cases - even though she is lacking any disposition to form a motive in line with her evaluative-focused belief. The theoretic plausibility of Eve's case - one beyond of whether it could be conceivable to think of Eve in a certain situation – rests basically on the plausibility of a psychological theory widely assumed. In terms of this theory, beliefs and desires are modally independent. Thus - and as far as moral beliefs conform a sub-set of beliefs and our philosphical theory is explanatory fruithful in non-moral cases - we should not accept IQ. No

metaphysical appeal to issues related to the conceivability of certain scenarios, Mele argues, is presuposed by this case againts IQ.

1.3 Shafer-Landau's version

Another suggestion available in support of an alternative sense of *possibility* to interpret premise (ii) in the argument sketched above has been recently defended by Russ Shafer-Landau [Shafer-Landau, R. 2003. p. 145 and 148-150]. In the terms proposed by Shafer-Landau we could understand premise (ii) without paying attention neither to whether certain situations are conceivable (and their importance for our conceptual uses) nor to the nature or the functional role associated with our psychological types. Shafer-Landau's account has the minimal aim to establish the initial plausibility of premise (ii) [Shafer-Landau, R. 2003. p. 145]

But why should we accept such minimal constraint to establish the soundness of (ii)? In order to answer this question, let me explain what Shafer-Landau is *doing*. He is presenting, in essence, a set of situations where some agents in unusual circumstances lack a moral motive to act in accordance with their self-assumed moral opinions[60]. Shafer-Landau expects to show (by noting some consequences around these examples) that our apparently pervasive intuition to favour internalism is, first and foremost, based on a narrow understanding of the multiplicity of contexts in which we normally utter a moral opinion [Shafer-Landau, R. 2003. p. 147]. He writes, for instance:

> "I believe that the examples I am about to offer create good, although not conclusive, reasons to accept the claim that the motivation provided by moral judgments is *prima facie* rather than *pro-tanto*. There are reasons for thinking that moral judgments that are able, by themselves, to motivate in some contexts may entirely

---

[60] Shafer-Landau understands externalist's central thesis as saying that whether an agent's motives are aligned with his assumed moral opinion is a contingent matter, which depends on the *content* of his moral judgment and on certain *contextual factors*. These contextual factors could determine, in Shafer-Landau's sense, both the possibility that a motive is associated with a moral opinion and the overall prevalence of such moral motive in the determination of the agent's actions (or intentions). See Shafer-Landau, R. 2003. p. 151

lack their ordinary motivational power in unusual circumstances" Shafer-Landau, R. 2003. p. 148-149

Shafer-Landau introduces two cases to motivate his minimal case in favour of premise (ii). Each case supports a different aspect of his interpretation of the externalist's basic claim. He starts by presenting examples where the *content* of a moral opinion, *what is being explicitly affirmed by uttering a moral opinion*, determines the motivational import of the judgment – Shafer-Landau refers to cases of *futility*, cases of *anticipated cost*, and cases of *inverted commas uses* of moral language. Shafer-Landau assumes, for instance, that in cases where the moral demand is "so strenuous as to be unsatisfiable in particular cases (...) it is conceptually possible for an individual who holds this view to fail to be appropriately motivated" [Shafer-Landau, R. 2003. p. 149]. Shafer Landau notes that unless there is something special about cases of *moral futility* (cases where the action demanded is not satisfiable by the agent in this situation), we should offer an explanation of why we are tempted to explain such cases along internalists lines, i.e. as cases where a moral motivation is defeated by a non-moral motive (and not as cases where no motivation is available). Shafer-Landau extracts an imortant piece of meta-philosophical advice from these situations. In his favoured terms, as far as the *content* of our moral judgments is useful to predict whether our moral opinions are motivationally active, motivational internalism cannot be understood as an unqualified metaphysical claim. We need to appeal to particular situations to determine its truth.

After that, he focuses on the cases where *the particular psychological make-up of the agent* prevents moral opinion from having any motivational influence – cases of *listlessness* or cases of *inverted comas uses related with psychological disorders* [Shafer-Landau, R. 2003. 148-151]. Again, the important point of all these cases, or so Shafer-Landau purports to argue, is that if they were accepted as plausible then, even if we did not have an overall rejection of *internalism* as a metaphysical claim, we could have at least a sufficient basis to support a strong case to re-evaluate the intuitive appeal of *internalism*. From a methodological standpoint, although Shafer-Landau does not argue for an overall rejection of *internalism* - in the way Brink did, for instance – he suggests a shift of focus on the domain where the polemic is usually discussed. He presents, in sum, a modification to make possible a discussion more mundane of the motivational issues surrounding the polemic between internalists and externalists. In a certain way, the sense of *possibility* associated with any *conceptual* rejection

of internalism would be minimal. It would only require to show that in a significant range of *cases* where an agent is unmoved by a self-assumed moral opinion, not being able to form a motive must be easily explained either because the *content* of his moral judgment does not favour such thing or because certain *contextual* factors are cancelling the motivational effect of his moral commitment. But if we do not need to presuppose a defeated motive to explain what is happening in such cases, the motivational reliability of our moral opinions is *contingent* with respect to a set of external factors about us – the content of our moral opinions and our particular psychological make-up at the moment we utter a moral opinion [Shafer-Landau, R. 2003. p. 151]

When we put together those insights unveiled by Shafer-Landau and Mele, something becomes clear. We do not need to appeal to an unqualified sense of possibility to attack internalism. We merely need to focus on particular scenarios (Shafer-Landau) or to offer a theoretically-based explanation (Mele) to explain how *possibly* we can sometimes endorse a moral opinion without being motivationally affected by such opinion to any extent.

2. Explanatory arguments against internalism

In the previous section, *motivational internalism* has been presented by attending to a core modal statement, which affirms that, as a matter of *conceptual necessity*, there is an internal sort of connection between the state expressed by an agent's acceptance for himself of a moral requirement and a motive to act in accordance with the content of his acceptance. As I have noted, this formulation implies that it would *not* be possible to fully endorse a moral opinion without being motivated to act, at some extent, as the content of our opinion requires. But some people have resisted this modal statement. They defend that it is possible that an agent endorses a moral opinion without his motives being influenced in any relevant way.

Internalism, however, could be rejected by other means that do not imply that its core modal statement is false. Among these alternative routes, Sigrún Svavarsdóttir has recently suggested a sort of *methodological* or *explanatory* option to minimize the force of *internalism*. Svavarsdóttir has used some insights extracted from the way our common explanations of observable phenomenon work to argue that we should prefer an external

model of motivation to conceptualize the dynamic character associated with our moral utterances [Svavarsdóttir S. 1999 and 2006].

In order to present Svavarsdóttir's *explanatory* case I need to quote an example that he uses that is crucial to grasp the precise level of description that Svavarsdóttir favours. He asks us to consider Patrick's case:

> **[Patrick's case]** "Virginia has put her social position at risk to help a politically persecuted stranger because she thinks that it is the right thing to do. Later she meets Patrick (…) Our morally committed heroine confronts Patrick, appealing first to his compassion for the victims. Patrick rather wearily tells her that he has no inclination to concern himself with the plight of strangers. Virginia then appeals to explicit moral considerations: in this case, helping the strangers is his moral obligation and a matter of fighting enormous injustice. Patrick readily declares that he agrees with her moral assessment, but nevertheless cannot be bothered to help. Virginia presses him further, arguing that the effort required is minimal and, given his position, will cost him close to nothing. Patrick responds that the cost is not really the issue, he just does not care to concern himself with such matters. Later he shows absolutely no sign of regret for either his remarks or his failure to help". Svavarsdóttir, S. 1999. p. 176-177.

Just after the quoted passage, Svavarsdóttir settles the stage in a certain direction [Svavarsdóttir, S. 1999. p. 177]. He assumes, first and foremost, that in considering Patrick's behaviour we are basically confronted with a set of observables phenomena, which can be described by taking into account Patrick's verbal utterances and behavioural events or actions. In a related way, Svavarsdóttir also assumes that we achieve an important degree of agreement between those interested in offering a theoretical statement of what is going on in such cases by remaining located at this descriptive level. In a minimal sense, Svavarsdóttir points out the fact that it is because we can agree on Patrick making certain utterances and performing or failing to perform a certain action (an action in line with the content expressed by the utterance) that we can offer a neutral basis to interpret or to understand (by means of a certain hypothesis) Patrick's overall performance.

Once we accept this minimal ground, Svavarsdóttir proposes a general argument around Patrick's case. This argument rests on a central methodological constraint that governs any

situation where two parties are trying to offer alternative hypothesis to explain a certain external phenomenon (in the just noted sense, i.e. by offering some processes or by pointing to some mental states as responsible of Patrick's overt performance). The argument proposed by Svavarsdóttir goes as follow:

(1) In understanding (in trying to figure out) the psychological states of Patrick we usually rely on present and past observations of Patrick's behaviour along two basic principles: (a) an epistemological principle of interpretative charity and (b) a general principle assuming the *stability* of certain dispositions associated with A's mental life [Svavarsdóttir, S. 1999. p. 177]

(2) Our epistemic situation in Patrick's case is an instance of more general cases where we try to figure Patrick's mental states by proposing a certain hypothesis about his mental life that could make sense of his overt behaviour.

(3) By (1) and (2), we might to rely on present and past observations of Patrick's overt behaviour – supplemented by (a) and (b) - to figure out his mental states.

(4) Once (3) is assumed, we observe that the current and past data we have about Patrick's behaviour in certain situations - Patrick uttering a moral requirement without behaving in line of his self-assumed moral assessment - are mutually consistent: Patrick has always been consistent in not caring about his moral utterances.

(5) Once we take into account all the data - and once they have been constrained by (a) and (b) – *an internalist theorist* should defend that the mental states expressed by Patrick in those cases where he is uttering a moral opinion without behaving in line of his moral assessment, should be understood either by referring to "some inkling of motivation to do what he judges morally right or good" (p. 176) or by referring to an *insincere* moral judgment. He (the internalist theorist) could never accept that Patrick's overt behaviour could be explained by "the hypothesis that Patrick is making a moral judgment but is not motivationally affected" [Svavarsdóttir, S. 2006. p. 165].

(6) Once we take all the data into account - and once such data have been constrained by (a) and (b) – *an externalist theorist* should affirm that the mental states that Patrick expresses when he is uttering a moral opinion without behaving in line of his moral assessment should be understood as referring to a state of acceptance of a *moral judgment* instantiated by Patrick plus *an absence of motivation* on Patrick's motivational side.

(7) A ***methodological principle of simplicity*** is regulating our explanatory practices**:** *"[W] hen there is a conflict of intuitions (among intelligent and sensible people) about what hypothesis are in the running as an explanation of some observable phenomenon, the burden of argument is on those who insist on a more restrictive class of explanation"* [see Svavarsdóttir, S. p. 180. 2006. p. 165]

(8) But if the methodological principle of simplicity is sound, then the option being favoured by internalism is more *restrictive,* form an *explanatory* standpoint*,* than the externalist's hypothesis

Let me explain further this argument. Let suppose that Patrick's overt behaviour could be encapsulated in three observational statements: 'Patrick uttered a sentence *X'*, ' Patrick was not motivated to behave as X was prescribing' and, finally, 'Patrick was not in regret after not being motivated to X'. Once we assume this observational basis, the internalist understands Patrick's overt behaviour either as (i) an expression of an insincere judgment or as (ii) marks pointing toward a hidden motivation to act in light of X. This proposal expresses a maximally restrictive interpretation of the facts because it does not refer to any of the observational statements assumed above. By contrary, the externalist understands Patrick's external behaviour as *transparent* in relation to the mental states expressed by Patrick's overt performance. So, Patrick's behaviour would simply express a moral judgment (X) along with a lack of motivation to act in line with X *as the first and second observational statements are assuming*. For this reason, Svavarsdóttir argues, externalism is *less* restrictive than internalism in relation to observable phenomena.

Internalists could attempt a rebuttal by assuming that the non-restrictive hypothesis, even if it is plausible by itself from a methodological standpoint, is in fact giving voice to a

conceptually incoherent position. They could defend that it might be conceptually required by our normal practices of moral ascription and moral appraisal to respect certain restrictions on the descriptive possibilities open when we describe cases like the one that Patrick exemplifies. In sum, they could argue that it is entirely justified, at least from our common practices, to reject an externalist's rendition to Patrick's case.

But Svavarsdóttir points toward a pervasive intuition. Although his example is focused on certain pieces of behaviour, it is a local consequence of a more general explanatory practice that we surely stand for. In terms of it, we regulate any explanatory ascription, in any context (even an imaginary one), in terms of *simplicity*, as premise (7) assumes. If this was accepted, internalists should offer an explanation of why in the local case exemplified by Patrick it would be acceptable to ignore such principle of simplicity when we try to explain Patrick's behaviour. If an explanation is not offered then we should accept that there are some cases where we would be better situated to explain parsimoniously certain human actions by taking an *external* framework on the nature of moral motivation [Svavarsdóttir, S. 1999. p. 181]

3. Securing internalism?

I have presented *internalism* as the philosophical translation of an intuitive thesis about moral motivation. Despite its initial directness, *internalism* is not free of problems and in the previous sections I have sketched a reduced battery of critics aimed to provoke a certain amount of scepticism among those that find the reconstruction that internalism offers of our intuitions of moral motivation a plausible achievement. However, internalists have not been impressed by these arguments. And this is a comprehensible thing. If internalists renounced to locate the meta-ethical debate into the motivational domain – they would dismiss a very attractive theoretical tool to overcome the *impasse* surrounding *Meta-ethics*. As I illustrated in chapter 3, many philosophers, from different meta-ethical sides, have appealed to argumentative structures for which internalism was an essential component to minimize such *impasse*. In this assumed framework, to reject *internalism* would be equivalent to renounce to a central component of some knockdown attempts to reinitiate *Meta-ethics*. For this reason, if I am right about this general methodological point, internalists should try further to secure

their favoured intuition about motivation, keeping alive the motivational path to dissolve the *impasse* referred above [McNaughton, D. 1988. p. 136-137. Jay Wallace, R. 2006b. p. 191-193].

3.1. Improving the conceptual strategy

Internalism has an immediate option open to resist the challenge presented by an externalist stance on moral motivation. The strategy works by merely acknowledging that, although *conceivable*, externalist scenarios do not express genuine instances of conceptual competence in applying or using basic moral terms. In the section to come I will be concerned with this particular strategy to reject externalism. Let me introduce the strategy first by focusing on the work of Hare [Hare, R. M. 1952. p. 165-172. Hare, R. M. 1981. p. 183-185].

3.1.1 Hare's version of the conceptual strategy

Among the conceptual-based strategies to defend *internalism* against the variety of criticisms I have just presented, Hare's proposal is a good starting point to contextualize the debate. As I defended in chapter 2, one of Hare's basic goals was to determine the importance of descriptive information in a meta-ethical account of the status of our evaluative thoughts [Hare, R. M. 1952. p. 111-122. 1981. p. 87-88]. To put in simple terms, Hare proposed a framework (*universal prescriptivism*) that accommodates a descriptive intuition strongly defended by moral realists – either of non-naturalist or naturalist tendency – without renouncing to the action-guiding character that is linked to our moral opinions [Hare, R. M. 1952. p. 134-136. 1981. p. 20-21].

In Hare's sense, the importance of descriptive features does not consist either in how they denote a realm of *moral* facts, nor in how they refer to a realm of *natural* facts. Instead, Hare assumed that we should understand the relevance of the descriptive features involved in our moral evaluations as enabling the rational construct of a feature even more basic of our moral thinking, its *prescriptive character* [Hare, R. M. 1952. p. 1, 46-47, 79, 155-158. 1981]. Hare argued that deliberative agents are committed to accept, by their implicit commitment with a formal requisite of *universalizability,* that any set of non-evaluative or descriptive features

used to evaluate a certain item should be able to ground a similar evaluative judgment in any similar and relevant situation [Hare, R. M. 1952. p. 134-135. 1963. p. 30-31. 1981. p. 21].

Hare noted that, although several people[61] had located the function of our moral utterances in a practical or cooperative context[62], the right way to understand the practicality that underlies our moral opinions should establish a bridge between the descriptive and the action-guiding character of moral opinions along the ideal of *universality* – an ideal that Hare understood as specifying the *type* of non-moral features that are governed by the previous requisite of *universalizability* [Hare, R. M. 1952. p. 129. Hare, 1963. p. 30 and 38-39]. A moral judgment could guide an agent's action in accord with the ideal of universalizability by appealing to reasons of a different kind than those adduced to justify speech acts that command or threaten someone to do something. Hare argued that the set of descriptive features on which moral statements are usually based would justify their utterance only if they were independent of the contingencies of the person who is uttering the statement (universality). But once this general framework is stated, how did Hare connect it with the motivational problem we are concerned with?

Hare's remarks about moral motivation and amorality are well known at this stage. In cases of amorality, Hare argued, the agent is surely using a moral term in an *inverted commas sense.* In inverted commas uses, an agent applies a certain moral term to a given situation to describe how a group of people use such moral term. The agent, rather than endorsing such use, is describing a certain standard (by uttering a moral opinion in the circumstances in which it would be right do so) [Hare, R. M. 1952. Especially at page 124 for an *evaluative*-focused example and at page 164 for a *deontic* inverted use. See also Hare, R. M. 1963. p. 74-75. Hare, R. H. 1981. p. 24 and 58].

Once the relative centrality of *inverted commas uses* of moral terms is noted, Hare use this very category to explain what is going on in the cases where an agent's moral utterance is not connected with his motives. Hare noted that in such cases the agent is not expressing a genuinely moral speech-act, but rather exemplifying a type of speech-act by means of which

---

[61] See Hare, R. M. 1952. p. 12-13. He is referring to Stevenson, Ayer and Carnap, who defend that the practical function of moral language can be reduced either to a commendatory (Stevenson) or emotive (Ayer) feeling aimed to causally affect other's behaviour or to move to action in the speaker's side.

[62] So, in the opening lines of Hare's first book, he states this intuition in an extremely elegant way: "If we were to ask of a person 'What are his moral principles?' the way in which we could be most sure of a true answer could be by studying what he did" And a bit later he adds: "The reasons why actions are in a peculiar way revelatory of moral principles is that the function of moral principles is to guide conduct. The language of morals is one sort of prescriptive language". Hare, R. M. 1952. p. 1.

he (the agent) *describes* how certain people behave in terms of an accepted standard – which implies, surely, a negative attitude toward such type of standard. Hare wrote:

> "We noticed that it is possible for people who have acquired very stable standards of values to come to treat value-judgments more and more as purely descriptive, and to let their evaluative force get weaker (…) Thus is possible to say that 'You ought to go and call on the So-and-sos' meaning by it no value judgment at all, but simply the descriptive judgment that such action is required in order to conform to an standard which people in general, or certain kind of people not specified but well understood, accept " Hare, R. M. 1952. p. 16

The essential point noted by Hare was that there is no logical incoherence in an agent that believes that some act is wrong and, at the same time, is not being motivated by his own judgment. We should be able to dismiss the apparent conceptual coherence that underlies these cases of amorality (Brink) by claiming that they do not really instantiate *genuine uses* of evaluative terms at all. They all exemplify *inverted commas uses* that exploit the descriptive side of our moral concepts.

3.1.2 Problems for Hare's version

Hare's argument is subject to a powerful and direct reply, one pursued by David Brink and, more recently, by Sigrún Svavarsdóttir, S. [Brink, D. O. 1989. Svavarsdóttir, S. 1999]. The reply goes as follow. Let us suppose that we accept Hare's solution. According to it, any agent who remains unmoved by a self-referred moral opinion would instantiate, as a matter of conceptual necessity, an *inverted commas use*. But to accept Hare's solution we should assume a conditional statement, one saying that

(C) If any deviant moral use (DM) can be explained by appealing to an inverted commas use (IC) then it should always be possible to identify a moral standard that would enable us to equate DM to IC.

Once assumed **C**, let me bring to your attention a possible scenario, which effective plausibility has been forcefully defended by Sigrún Svavarsdóttir. In this imagined scenario two things are assumed to be true:

(i)   An agent (A) is not motivated to ϕ at t_1 even though at t_0 he judged that ϕ was the right thing to do

and

(ii)   A's moral judgment - *that ϕ is the right thing to do* - is neither referring to a moral standard shared by a minority group into A's own community nor expressing the views endorsed by members of other community.

Evidently, in the scenario depicted by (i) and (ii) A is adopting a quite peculiar moral perspective. But even so is accepted, we could tentatively accept that he is uttering a *moral* sentence. In such scenario, either he is judging that ϕ is the thing to do from an *omniscient perspective* (one from where he would grasp a certain moral truth in a privileged sort of way) or he is morally disposed toward ϕ from a *maverick stance* (one that enables him to endorse a completely idiosyncratic moral opinion about the status of ϕ) [Svavarsdóttir, S. 1999. p. 176].

But once this scenario is presented – and keeping in mind the different perspectives that A could exemplify there - what does it say about Hare's overall strategy? The response to this question is quite direct, or so Svavarsdóttir argues. If possible, moral judgments that are endorsed from a *maverick perspective* and cases of *evaluative omniscience* exemplify scenarios where an agent lacks any grasp of common standards. And because of this lack of common standards both cases would not satisfy the condition fixed by C (a condition, remember, implied by Hare's favoured explanation of what is going on in cases of amorality - *inverted commas account*). So, if we accepted that in the cases favoured by Svavarsdóttir A is unmotivated to act in line with his self-assumed moral opinion, then we should *not* assume (as

Hare did) that A is *not* making a genuine moral judgment. As far as in neither of the cases favoured by Svavarsdóttir A is having a grasp of any convention support for his detached stance toward φ (as C is requiring) we cannot assume, as Hare did, that A is not making a genuine moral judgment. In sum, by accepting that in certain situations the utterance of a moral term cannot be explained by appealing to an *inverted commas* model (because the uses stressed by this model are constitutively grounded on shared moral standards and in such situations the agent uttering the sentence is not contemplating any standard at all), some philosophers have questioned one of the most important arguments to defend motivational internalism.

3.2. A methodological rejoinder from internalism

I noted before that *internalism* does not only face *conceptual* problems about its preferred formulation, i.e. problems related to the conditions of applicability associated with our basic moral concepts. Besides, the explanatory ideal that internalism favours could be in deep disagreement with some pivotal intuitions about how *parsimony*, *economy,* or simplicity constraint any proposed explanation, in any context. In the next section I will be concerned with a sort of explanatory rebuttal coming from *internalism*. In terms of this movement, *externalism* can make sense of why an agent successfully obeys a moral requirement only by presupposing a dubious image of the type of character involved in genuine moral motivation. In what follows I will attend to this general intuition.

3.2.1 Smith's *moral fetishism argument*

Smith starts by acknowledging that the two main views on moral motivation (*internalism* and *externalism*) should agree on the motivational force of our self-assumed moral opinions. So, although *internalism* and *externalism* disagree on how the intimate connection between our moral opinion and our motives should be explained, both theories might agree on a minimal point: A motivation to act in favour of φ follows reliably from a moral opinion that says that φ is the right thing to do - at least *in a moral or strong-willed person*. I will refer to this shared intuition as (Ct). Once this intuitive agreement around (Ct) is acknowledged, Smith will assume, as I did at the beginning of chapter 3, that both theories defending two opposite

theoretical frameworks in order to explain the motivational reliability encapsulated by (Ct) [Smith, M. 1994. p. 71-72. Dreier, J. 2006/2000. p. 548].

An internal explanation of (Ct) would claim, in essence, that the conjunction of the *content* of a moral opinion and the force of certain *psychological state* explains the motivational reliability of our moral statements. An *externalist explanation* of (Ct) would assume, in turn, that the motivational reliability of our moral opinions (being a moral opinion the philosophical composite extracted from a certain *attitude* focused toward a certain *content*) is explained by an additional psychological state other than the one expressed by the moral opinion itself [McNaughton, D. 1988. p. 134. Smith, M.1994. p. 73-74]

Once we assume this general framework, Smith's case in favour of the *explanatory* prevalence of internalists theories works, in essence, by assuming that a very general capacity underlies our evaluative thoughts in deliberative contexts, i.e. ones where an agent must weight different options to reach an overall judgment about what to do. Smith assumes that in these contexts we are able to exemplify a general *capacity* by means of which our motivational system could track the content of our evaluative beliefs in determining our motives. This capacity suggests, evidently, a basic and quite general *constraint* governing our deliberative processes, one that could say something like *in any deliberative context*, *form your motives in light of your evaluative deliberation*. I will refer to such constraint, following James Dreier, as *the tracking condition* [Smith, M. 2004. p. 273-274 and Dreier, J. 2006/2000. p. 549].

As I have just said, *the tracking condition* fixes a very intuitive *constraint* on our deliberative practices [Dreier, J. 2006/2000. p. 550]. For instance, when I have good evidence against an evaluative assertion, the thing to do, at least if I am respecting the constraint I referred above, is to lack any motive derived from my previous (and wrong) evaluative belief (and to lack it in a non-deviant way, in virtue of my own recognition that my previously assumed claim is not supported by evidence). But *the tracking condition* also exemplifies a psychological phenomenon in need of explanation. Surely, any proposed explanation of *the tracking condition* should take into account not only the nature of the *reliability* by itself - in Smith's own words: "(…) why I change my motivation when I change my judgment" [Smith, M. 1994. p. 74] - but also the type of psychological mechanisms that support such reliability. Thus, besides of the constraint saying that our motives ought to be reliably connected to our evaluative appraisals, we should consider how such constraint is supported by certain

psychological mechanism. In sum, if we were explaining the nature of the *tracking condition* we should be focused on a single question: What sort of psychological capacity or disposition could fulfil the functional role that is suggested by the *tracking condition* in cases of evaluative or moral deliberation?

In order to say something about this, let me consider first a typical case of moral deliberation. Suppose, for example, that A judges at t_1 that X is the right course of action. Let us say that A thinks that to attend to a certain demonstration against a policy enacted by his own government is the right thing to do. However, after a deliberative intercourse with B, A rejects his previous judgment, thinking at t_2 that Y rather than X is the right thing to do – when 'Y' refers to any action-type aimed to support the policy of the government. Accordingly to *the tracking condition*, A's motives should be aligned with his own evaluative judgments. Traditionally, in the terms favoured by externalism, A's motivation to act in line of Y could be secured by an external desire, which content is 'to do what is the right thing to do' [Dreier, J. 2006/2000. p. 548]:

> (*External desire*) A desire to do what is right

But the desire favoured by *externalism* to secure *the tracking condition* is ambiguous between two possible readings. The external desire could be understood either as implying

> (*de re* **interpretation**) From each thing that is in fact right, A desires to do it

or as implying something like

> (*de dicto* **interpretation**) A desires to do whatever is right

Take the first possible interpretation, which implies a *de re* reading of the external desire. In terms of such interpretation, when A judges at t_0 that Y rather than X is the right thing to do, A's motivation for Y at t_2 will be enacted by his taking into account at t_1 the normative status of Y itself - surely trough an attitude focused on certain evaluative features of Y. The desire we attribute to A is understood along *de re* model of attribution precisely in virtue of how the features associated with Y are responsible to enact by themselves the desire – in our

example the desire for Y. In essence, although in a *de re* interpretation the external component remains unmoved, it is constitutively linked to certain features of the evaluated item[63].

But there is a problem with this interpretation of the external desire, or so Smith argues. And it is an important one: if the additional component proposed by externalism to explain the motivational reliability of our moral opinions is understood along a *de re* reading, then *the tracking condition* cannot be secured. Because if we accept a *de re* reading how could we explain changes in A's motives by merely appealing to the fact that A is judging or believing that Y *itself* is the right thing to do. As Smith notes:

> "(…) the mere fact that I have found reason to change my
> judgment gives me no reason to change my motive (…)
> Having a *non-derivative concern* for libertarian values
> while judging it right to vote for the libertarian is thus not
> what makes me a good person. For it cannot explain why
> I change my motivation when I change my judgment"
> Smith, M. 1994. p. 74. Cursive mine.

But, as I noted before, externalists could embrace a *de dicto* interpretation of the external desire instead. In this interpretation, when A judges that Y is the right thing to do, the motivational reliability associated with A's moral opinion is secured by the intervention of a non-derivative desire, which content is 'to do whatever is right to do'. In these terms, when a *de dicto* desire is conjoined with an evaluative belief which content is 'Y is the right thing to do' both states cause a *derivative desire* in A for Y-ing[64].

Once these distinctions are assumed, Smith makes his central movement in support of an explanatory argument for *internalism.* Let us suppose, he argues, that we accepted that to explain how moral opinions and motivations are connected along the tracking condition we might assume a *de dicto* desire which content is 'to desire to do whatever is right to do'. Let

---

[63] A far as a derivative desire is a variety of desire – being required as such to secure A's motivation.

[64] The reliability of the connection between judging that Y is the right thing to do and effectively doing Y is secured in a *de dicto*-based schema by assuming that when I decide to change my preference from X to Y, what I am losing is a *derived* desire for X, which is functionally dependent upon a non-derivative or a standing desire with the content of 'to do whatever is right to do' plus a certain evaluative belief about an action-type (X or Y). Once the evaluative belief is changed at t_2 (in this case in favour of Y) A will instantiate a new derivative desire for Y as consequence of my standing desire 'to do whatever is right to do' plus the evaluative belief affirming the rightness of Y.

us suppose that it is always operative in enabling the action-guiding role associated with our moral opinions. Now, let us consider a set of well-established intuitions about the type of motivation that underlies the cases where an agent is motivated to do what he has judged as right. Suppose, for example, that we accept that " (…) [g]ood people care non-derivatively about honesty, the weal and woe of their children and friends, the well-being of their fellows, people getting what they deserve (…)" [Smith, M. 1996. p. 75]. How could externalists put together both points? How could they put together the fact that moral motivation is enabled by the presence of a *non-derivative* desire to do 'whatever is the right thing to do' and widely assumed intuition about our character traits, which say that virtuous agents might be able to care non-derivatively about certain experiences, persons, and things - rather than to care about the very fact that such items are worth-caring?

But if non-derivative concerns shape our views about moral character and virtue, then externalism would face an important dilemma. In terms of it, if *externalism* were understood along a *de dicto* interpretation it would be very complicated to fit the motivational profile favoured by this particular interpretation with some intuitions about the type of tracking-relation that virtuous agents exemplify. On the other hand, if *externalism* were understood along a *de re* interpretation, then the motivational reliability of our moral judgment could not be coherently explained. A *de re*-based schema appeals to certain evaluative beliefs, which focus on certain features of the evaluated item, we cannot explain how the non-derivative desire is brought into existence. As far as a *de re*-based schema goes against our better philosophical psychology we should we endorse the first horn of the dilemma. And this is *the fetishism argument* [Wallace, R. J. 2006a. p. 329].

If Smith is right about its consequences, neither of the possible interpretations of the desire posited by *externalism* would secure a plausible psychological profile to explain how we can fulfil *the tracking condition*. And sadly, we would leave the motivational reliability of our moral judgments fully unexplained. But how plausible is this line of rejection of externalism? How plausible, in sum, is *Smith's moral fetishism argument*?

3.2.2. Problems for Smith's moral fetishism argument

It is important to delimitate precisely the scope of the argument just noted. The argument is *not* saying that cases of moral motivation cannot be explained conceivably along externalist's

assumptions. The argument is claiming rather that if we accepted *externalism* as a general explanatory framework then, no matter what version of externalism we preferred, we would be doomed to embrace a very counterintuitive claim. Either we might argue against the received view in philosophical psychology (*de re* interpretation) or – once we express our commitments with this orthodoxy[65] - we might ignore an intuition widely assumed about particular instances of virtuous caring or concern (*de dicto* interpretation). At the end of the day (and given our acceptance of the orthodoxy in philosophical psychology) neither of the available routes would offer a sensible explanation of how our motives are tracking our evaluations in a non-deviant way.

As I see the debate at this stage – and keeping in mind, again, our acceptance of the standard account of motivation - there are two available options for externalism. First, externalists could question the relevance Smith assigns to a certain interpretation of *the tracking-condition*. Second, externalists could simply argue, in a quite fearless way, that our intuitions pertaining to how a virtuous agent is responsive to his self-assumed moral opinions are unfounded or, to say the less, extremely narrow in light of the variety of reactions that compose a complex moral perspective [Smith, M. 1994. p. 76].

I am not going to argue directly for second option here. In turn, I wish to sketch an alternative that takes advantage from the first available route. In essence, I would like to refer to a suggestion that notes the possibility of conceiving the externalist perspective, i.e. the psychological scaffolding required to support it, through a lens less *opaque*. Some externalists have argued, for instance, that a quite plausible route to overcome Smith's fetishism objection would simply note that moral agents are usually equipped with a direct or *transparent* kind of disposition to be motivated to do whatever thing is judged by them as right [Broome, J. 1997. Scanlon, T.1998. Chapter 1. Dreier, J. 2006/2000. p. 551].

Robert Jay Wallace has recently illustrated this model by locating it into a wider sort of inquiry focused on the nature of those psychological capacities that enable us to act in light of our judgements about what we ought *rationally* to do [Wallace, R. J. 2006/1999]. In Wallace's hands, the just proposed externalist model should appeal to a *basal disposition* to secure a non-deviant tracking of our normative thoughts by our motives [see Wedgwood, R. 2007. p. 27-28 for a discussion on the nature of this disposition]. Wallace importantly notes that such *basal disposition* is not mediated by any belief about the rationality of doing a

---

[65] This is Smith's own case.

certain action. In Wallace's sense, a basal disposition could secure directly our motivation to do what we judge to be rational without requiring the presence of any non-derivative desire:

> "One's rationality is displayed in one's having a tendency to change one's view in ways that, as a matter of fact, enhance coherence; but no causal role is played by agent's own acknowledgement that transitions of this kind are requirements of reasons (…) rational agents are guided by rational requirements. They are not guided by their acknowledgment of inferences and actions as rationally required" Wallace, R. J. 2006/1999. p. 53

In essence, if we accept Wallace's remarks, it could not be true that any external account of the *tracking condition* must be understood necessarily as referring to a substantive concern - as *de dicto* interpretation presented by Smith assumed. Rather, the non-cognitive or external component could simply refer to a disposition directly triggered by our normative thought in a *transparent* sort of way. If so, moral agents could be non-derivatively motivated to act in line with their normative judgments without presupposing any degree of *fetishism* on their performances.

4. Looking back and looking forward

Let's recap. In the last two chapters I have been concerned with an argumentative framework that aims to offer a direct route to resolve the *impasse* surrounding *Meta-ethics* (chapter 2, section 4). The structure of this framework can be encapsulated by three general statements:

> *(1) A conceptual claim about the motivational import of our moral opinions:* Moral opinions are reliably connected with our motives to act.

> *(2) An accepted theory explaining human motivation:* Any action of A could be explained by referring to A's psychological factors or dispositions.

> *(3) A methodological gambit:* If (1) is conceptually sound, then any meta-ethical approach asking for the *psychological* nature of moral opinions should take into account the sort of findings proper of (2).

In my terms, if these assumptions were sound we could arrive, by means of a more precise interpretation of each one of the statements contained in (1)-(3), to a knockdown answer about the type of mental state expressed by our moral utterances. I presented several of these arguments at the beginning of chapter 3. But beyond the merit of these arguments I was concerned with a central question: Could we really arrive to a direct answer about the type of mental state expressed by our moral opinions by appealing to an structure such as (1)-(3)?

In relation with this question I have defended a particular style of answer, one located at a meta-methodological level. More precisely, I have favoured an approach to the previous question focused on whether a certain amount of disagreement around one of the assumption contained in (1)-(3) could justify our rejection of *any* particular strategy that, by appealing to (1)-(3), were aimed to overcome the *impasse* in *Meta-ethics*. In my sense, at least one among the assumptions contained in (1)-(3) is deeply problematic in virtue of the disagreement it incites. In this sense, all along these chapters I have used premise (1) to illustrate my strategy.

In order to motivate my case around premise (1) I have presented what I took as a standard or philosophically qualified version of the assumption underlying (1) – *internalism* (chapter 3). After that, I have developed a battery of critics focused on my qualified version of *internalism*. I have assumed that the combined effect of these criticisms which stem from a different philosophical view on the nature of our moral motivation (externalism) and the further rebuttals endorsed by some internalists would suggest an scenario where any argumentative appeal to the disputed assumption - (1) - should be considered as groundless or at least as deeply underdetermined. Therefore, the gains of embracing any strategy supported by the structure going from (1) to (3) to overcome the impasse surrounding *Meta-ethics* are less than optimal at the end of the day.

But because of the fact that approaches to *Meta-ethics* based on motivation are less than optimal routes to arrive to straight meta-ethical answers, some philosophers have argued for an alternative framework. They have appealed to a normative-based view of morality to determine - through their acceptance of a different general argument - the type of mental state that underlies our moral utterances. In the next two chapters I will present this general

argument, and will put it under the same type of criticism than I used against the motivational route.

Chapter 5
Meta-ethics and Normative Reasons

1. Introduction

2. Reason-based issues in recent Meta-ethics; A brief overview

3. *Internalism about reasons*: A preliminary formulation

4. An improved version of Internalism about reasons

4.1 The motivational set

4.2 The deliberative processes

5. Two basic arguments supporting internalism

6. Looking forward

Chapter 5

1. Introduction

Until now (chapters 3 and 4), I have been concerned with a single question:

> (Q $_{motivation}$) How could we obtain some insight about the psychological nature of our moral opinions by reflecting on the *motivating reasons* offered by morality?

I have assumed that this question, if framed by a certain argumentative structure, is useless to get a direct answer about the psychological nature of our moral opinions. In the following two chapters - and pressing a bit further on the hopes to reach a positive conclusion about the psychological status of our moral opinions - I will ask instead:

> (Q $_{normative}$) How could we obtain insight about the psychological nature of our moral opinions by reflecting on the *normative reasons* offered by morality?

In the same way that I assumed before an unified framework to do *Meta-ethics* by departing from motivational issues I will argue now that if *Meta-ethics* is intended as a more general reflection on the nature of the type of justification offered by morality, i.e. on the nature of the moral reasons, we should focus on a common argument [Foot, P. 2002/1972. p. 160-161. Brink, D. O.1987 p. 52. Smith, M. 1994. 61-65 and Shafer-Landau, R. 2003. p. 170]:

> (i)   If A is under a moral obligation to ϕ then A surely has a good reason to ϕ [*Conceptual claim about morality*][66]

---

[66] Until further notice (see Chapter 7 for a more detailed account) the sense of *rationalism* I am presupposing here is close to one recently stressed by Shafer-Landau and Cuneo. They write: "Moral rationalism asserts that there is an entailment, a necessary connection, between moral duties and excellent reasons for action. On this view there is always very good reasons to do as morality says (…) Moral rationalism divide on the issue of the strength of moral reasons. Some believe that morality always supplies overriding reasons – reasons that uniformly take precedence over competing considerations. Others believe that morality, though always providing strong reasons for action, can sometimes be overridden in unusual, extreme circumstances" Shafer-Landau, R. Cuneo, T. 2007. p. 281-282. Or consider Jay Wallace's statement: "(…) moral considerations at least purport to have normative significance. They present themselves to us as reasons for action, in the basic normative sense of being considerations that count for or against courses of action that are open to us" Wallace, R. J. 2006b. p. 186.

(ii)   A has a good reason to ϕ only if A can be motivated to ϕ f [*Substantive claim about the status of reasons*]

(iii)   A can be motivated to ϕ only if certain psychological conditions are the case [*Substantive claim about the nature of human action*]

Therefore – by (i)-(iii) A is under a moral obligation to Φ only if certain psychological conditions are the case

Importantly, the chain of statements presented above does *not* compound an argument by itself. In my favoured terms (see chapter 3, section 2) the *argumentative force* of the structure just noted rests on the particular content we assign to each of the assumptions that conform it. Its pervasiveness will depend, so to speak, on the particular thesis we choose to unfold the status of our normative reasons or on the particular theory we embrace to explain human action. Depending of the relative plausibility of these more particular choices, we could achieve a more or less plausible argument in support of a given meta-ethical position. Keeping in mind this general point, the previous schema has been sometimes interpreted as giving voice to a more concrete argument. In terms of it:

(1)   Necessary, to be under a moral obligation to ϕ is to have a good reason to ϕ [*Rationalism*]

(2)   Necessary, if A has a good reason to ϕ then A can be motivated to ϕ [*Reason internalism*]

(3)   Necessary, if A is motivated to ϕ, then A must either desire to ϕ or desire to φ and believe that by φ-ing she will ϕ [*Humean theory of motivation*]

Therefore – by (1) to (3) –A is under a moral obligation to ϕ only if either she desires to ϕ or she desires to φ and believes that by ϕ-ing she will φ.

In the next two chapters I will focus on the particular arguments that fill the *structure* offered by (i) to (iii) – a structure that I will refer to as E hereafter. My concern, in a certain

way, is animated by a single intuition about the status of E (and about the prospects of any *argument* derived from E). In essence, I believe that *any* particular argument that fills the space provided by E is not going to be well supported because of the simple fact that *at least one of the claims* that E assumes cannot be formulated without avoiding disagreement. In methodological terms favoured by my approach, we do *not* need to prove that *any* particular interpretation of E is false to question the plausibility of E to offer a solid framework to restart *Meta-ethics*. Instead, it would be sufficient to prove that at least *one* of the assumptions contained in E is *essentially underdescribed*. As I noted in chapter 3, an assumption is *essentially underdescribed* if it falls under the following description:

> **[U]** An assumption *p* located into a general argumentative structure E
> is *underdescribed* only if we do not have an instance of *p (p')* that is
> able to motivate a agreement in support of any particular
> argumentative structure (E') - one where a positive conclusion is
> entailed from *p'* along with other premises (*q', r', ... n*)

Once we keep U in mind, I will suggest that the most disputed assumption in the structure I sketched above is the one focused on the nature of our normative reasons. I have referred to this assumption as (ii) - or as premise 2 in the real argument - and it is going to be my focus in what follows. I will introduce a qualified version of (ii) in the current chapter, explaining later (chapter 6) why even this qualified version could exemplify a profound disagreement around those meta-ethicist that stress the importance of a normative-based route to overcome the *impasse* surrounding meta-ethics (Chapter 2).

2. Reason-based issues in *Meta-ethics:* A brief overview

My current focus is the following argument:

(1)   Necessary, to be under a moral obligation to ɸ is to have a good reason to ɸ [*Rationalism*]

(2)   Necessary, A has a good reason to ɸ only if A can be motivated to ɸ [*Reason internalism*]

(3)   Necessary, if A is motivated to ɸ, then A must either desire to ɸ or desire to φ and believe that by φ-ing she will ɸ [*Humean theory of motivation*]

Therefore – by accepting (i) to (iv) above –A is under a moral obligation to ɸ only if either she desires to ɸ or she desire to φ and believes that by φ-ing she will ɸ.

If sound, the above argument could help us to establish the psychological nature of our moral opinions by previously accepting a particular story about the authority of our moral demands (*rationalism*), a certain view on the status of our normative reasons (*reasons internalism*) and, finally, an account focused on the nature of our motivating states (*humean theory of motivation*). But additionally, the argument could be very useful in order to facilitate a fast overview of the place of normative issues in recent meta-ethics.

In the current section I will be concerned with offering such fast-and-frugal overview. I will introduce the range of conceptual possibilities around E by asking previously about the soundness of the argument going from (1) to (3). Let me start by the ***first premise***:

(1) Necessary, to be under a moral obligation to ɸ is to have a good reason to ɸ [*Rationalism*]

(1) can be denied along several lines [Blackburn, S. 1998 and 2006a]. However, here I will choose a route that appeals to two opposite intuitions to undermine (1). The first intuition assumes that moral requirements are normative (require something from us) with full independence from our desires, preferences, or goals. They are *categorical* in the sense that they require that we act in a certain way, without taking into account our motivational profile to determine the force of the demanded action or choice [See Foot, P. 2002/1972. p. 160-161 for a classic discussion around this point and Joyce, R. 2002. p. 42-44 for a more recent one].

The second intuition says that normative reasons are based, in some way or another, on our desires, motives, goals or preferences. They require something of us in virtue of how it could advance our goals or preferences [Foot, P. 2002/1972. 158-159].

Now, let us suppose that in a certain context (F) we were *not* behaving as the moral requirement prescribes – and we were aware of that. If (1) were true, we would be behaving irrationally in F - as far it is *prima facie* irrational not to be responsive to an acknowledged reason [Raz, J. 2001b. p. 67]. However, neither the rationality nor the irrationality of acting against a given demand in a certain context is necessary determined by the nature or status of the source of the demand itself – as the second point stresses. Therefore, if even in the cases in which an agent is ignoring a moral demand he could be behaving rationally, then we should accept that, contrary to premise (1), moral demands do not give voice to reasons.

But, evidently, as far as this is a far-fetched objection, those who defend (1) could accept the criticism just noted without renouncing to establish a connection between moral demands and reasons. They could assume, for instance, that the unconditional sense that defines the *scope* or *applicability* of our moral demands already suggests a stronger sense of *normativity,* by means of which we could justify a given action by A even if it does not bring any goal or preference of A. They could argue that our moral demands express this additional *categoricity* when they track a certain set of evaluative features, ones not dependent upon our desires. But this movement, one deriving an unconditional sense of normativity from the unconditional scope of certain requirements, cannot work either. To dismiss such extension, Philippa Foot famously noted that in some cases an agent can be under a demand which *applicability* is not dependent on his own desires - *etiquette* - but which capacity to *justify* the agent's actions in terms of reasons depends entirely on how the action that the (unconditional) normative system demands would fulfil the agent's goals or preferences. Foot argued that if this were the case, those who defend an extension of the justificatory power of moral reasons in line with the unconditional applicability of moral norms, should explain why in the moral context this could be desired while in another cases (*etiquette*) the extension does not work at all (because the *scope* or *applicability* of the system of norms of etiquette does not define the justificatory *force* that they could offer in terms of reasons). But as far as this explanation is not facilitated, those who defend (1) should conclude, or so Foot assumed, that moral demands are not giving voice to reasons in this derived sense either. Thus, as soon as premise (1) is not well secured

the rest of the argument cannot work because it rests on the concept of normative or justificatory reason [Foot, P. 2002/1972. p. 160-161. Joyce, R. 2006. p. 61-64].

But we also can resist the argument going from (1) to (3) above by attacking its ***second premise***. Remember that it claimed that:

> (2) Necessary, A has a good reason to ϕ only if A could be motivated to ϕ [*Reason-internalism*]

As we noted before, some people defend that sometimes we can *justify* our doing ϕ even when no counterfactual connection can be established between our doing ϕ and the satisfaction of our informed desires or preferences. Although this claim encapsulates the core of externalism, there are in fact two available externalists routes to reject premise (2). In line with a *pure* version of *externalism*, the type of relation usually exemplified by reasons (one where a certain *fact* favours a certain attitude [Scanlon, T. 1998. p. 19. Dancy, J. 2000. Ch. 1]) would refer either to the commonsensical domain of facts (one on which the reason-relation supervenes) [Parfit, D. 2001. Cuneo, T. 2007] or to a further, non-natural domain of non-reducible, purely evaluative, facts [Parfit, D. 1997. 2001 and 2006. Scanlon, T. 1998. Shafer-Landau, R. 2003. Dancy, J. 2006 and Enoch, D. 2007].

Along with a *non-pure* version of normative externalism, normative facts about reasons, although not directly connected to our own desires or motives, might well be resting on *facts* about our own psychology, i.e. facts about the type of psychological states we should form if we were subject to certain set of basic requirements such as coherence, unity or informedness [Smith, M. 1994.2004/1997 and 2005. Railton, P. 2004/1986 and 2004/1986a. McNaughton, D. Rawling, P. 2004].

Evidently, once either version of externalism is assumed, the irrealist conclusion that derives from the general argument is severely debunked. Even if premise (1) were true, at the very moment that we accepted that our normative reasons are *not* dependent on our desires or motives, a cognitive route to unfold the psychological nature of our moral opinions becomes available. In terms of this route, moral opinions would express beliefs rather than desires – as the conclusion of the original argument assumes.

Finally, can also attack ***Premise (3)*** in the argument above, the one assuming that:

(3) Necessary, if A is motivated to ϕ, then A must either desire to ϕ or desire to φ and believe that by φ-ing she will ϕ [*Humean theory of motivation*]

Premise (3) is committed with what is sometimes called *the humean theory of motivation - HTM* [Smith, M. 1987 and 1994. Ch. 4. Shafer-Landau, R. 2003. p. 127-140]. The HTM has been extensively defended by the assumption of an *a priori* argument, which defends that:

(a) Having a motivating reason is having a goal

(b) Having a goal is to be in a state with a world-to-mind *direction of fit*[67]

(c) To be in a state with a world-to-mind direction of fit is to desire

Therefore, having motivating reasons is to desire

Leaving the details aside, the argument above is aimed to establish that certain type of pro-attitudes (desires) are constitutively or conceptually involved in our motivation. As Shafer-Landau notes "(…) that whatever an agent intentionally Φs, we can truly say of her that she wanted to Φ"[68] [Smith, M. 1994. p. 116-119. See also Shafer-Landau, R. 2003. p. 134-141 and Tenenbaum, S. 2006. p. 250-258].

We have several available routes to resist the appeal of the argument in favour of the HTM just noted [McDowell, J. 1998/1978. Althman, J. 1986. Schueler, G. F. 2003. Finlay, S. 2007]. Nevertheless, some people have endorsed a different path other than rejecting the above chain of statements in order to resist to the prevalence of the HTM. In terms of this more positive route, once we assume that sometimes we are motivated by some normative belief, i.e. one focused on certain actions whose evaluative properties are unrelated with any of our goals or preferences, any adequate theory of motivation should accommodate this commonsensical fact in order to delineate better the nature and possibilities that define our motivating states [Nagel, T. 1970.Wallace, R. J. 2006/1990].

Once it is assumed that a desire is *not* always conceptually required to constitute a *motivating reason* - because, in Nagel's famous expression, our desires certain are normally

---

[67] See Anscombe, E. M. 1963. p. 56
[68] Shafer-Landau, R. 2003. p. 138

motivated by evaluative cognitions - the basic meta-ethical consequence endorsed by (4) - *non-cognitivism* – will be seriously undermined. Even if (1) and (2) were granted, moral opinions would be expressing beliefs rather than desires – because normative beliefs, so to speak, could also motivate by themselves without the assistance of any desire[69].

I have offered this minimal map as a preliminary step for the substantive work to be done in the current chapter and the following one. As I noted before, once we are aware of the routes available to resist the general argument going/that goes from (1) to (3), we should focus on a controversial premise of the argument and try to prove that there is a deep and most likely irresolvable disagreement, about the best way to formulate the premise in question. In what follows, I will focus on the premise about the status of our normative reasons (2). By acknowledging its disputed status I hope to highlight some interesting features associated with our concept of *reason* in a *justificatory* sense. After that (chapter 6) I will suggest that *any* attempt to find a direct answer to the question about the type of psychological states expressed by our moral judgments, i.e. any attempt supported by an argument in line with (i) to (iii), is misconceived because of a simple fact that at least one of the premises that have to be included in any instance of this argument-type (premise 2) will be essentially underdescribed.

3. Internalism about Reasons. A preliminary formulation

I have tentatively assumed (section 1) that we could solve the *impasse* surrounding *Meta-ethics* by accepting an argumentative framework that assumes certain points about the status of our normative reasons and their link with our moral demands. Remember that on a particular reading of (ii), it is interpreted as saying that

> (2) Necessary, A has a good reason to ϕ only if she could be motivated to ϕ [*Reason-internalism*]

---

[69] Although see Smith, M. 1995. 1997. 2001. 2003 and 2005 for a somewhat less direct version of motivational cognitivism, which assumes the fundamental role played by a *disposition to be coherent*.

Clearly, the most discussed version of (2) is the one introduced by Bernard Williams in his well-known paper 'Internal and External Reasons' [Williams, B. 1981]. An important concern Williams had in mind when he introduced his favoured version of premise (2) –one that I will refer to hereafter as WI - was clearly *explanatory*[70]. Williams acknowledged several times the centrality of this explanatory dimension to our common understanding of the kind of warrant offered by normative reasons [Williams, B. 1981. p. 102 and 106 and 1995. p. 39]. He assumed, in essence, that it should be possible to act on the basis of the good reasons that we have in a given context. This fact would make an explanation of our actions by appeal to these justifying reasons inmediatedly available.

Once this explanatory ideal was assumed, Williams asked if we could find some account of what reasons are, i.e. some view about the type of warrant required to ascribe a reason-predicate to the relation between an agent and a certain action, disposition or state (see footnote 4). To answer this general question, Williams rejected an obvious formula aimed to unfold the type of warrant and explanatory import that surround our reasons-statements. Williams referred to this failed attempt as *the sub-humean model of reasons*. In terms of a *sub-humean* account of premise (2):

> **(W1)** Necessary, if A has a normative reason to ϕ then A desires to φ and A assumes that he will attain φ by ϕ-ing

According to W1, an agent has a normative reason to ϕ only if she has a certain desire "whose satisfaction will be served by his ϕ-ing" [Williams, B. 1981. p. 101]. Williams rejects W1 adducing that it is "too simple" [Williams, B. 1981. p. 102. See Smith, M. 2004a for a more nuanced account of instrumentalism]. In Williams's sense, the basic failure of W1 is that it does not offer, as it stands, a sensible stance to grasp a basic feature linked to our normative reasons. In terms of this feature, the fact that an agent has a normative reason to ϕ

---

[70]   There is disagreement about this point. For instance, sometimes WI is presented as a thesis about the **metaphysical** grounds for justifying or good reasons. WI would affirm here that our normative reasons are *psychological facts* about the motives that we could have under certain conditions. See Scanlon, T. 1998. 364-365 and Dancy, J. 2000. p. 16. WI is presented another times as a **conceptual** claim, which unfolds the conceptual structure that underlies the uses that exploit the concept/s of *reasons*. Finally, WI is sometimes rendered as a thesis pointing to the **content** of our normative thoughts. Here, WI would be concerned with the type of processes we exemplify, in a rough way, when thinking in terms of reasons. See Williams, B. 1981. p. 106. Gibbard, A. 2003. p. 10-20. Pettit, P. 2006. Keeping all these possible readings in mind, I will assume hereafter that WI is better understood as simply referring to a thesis about the type of warrant required to ascribe a reason-predicate to an agent.

is usually connected to some *epistemic* or *informational* constraints focused on agent's motivational set. As far as W1 makes reference to an unqualified motivational profile, W1 cannot be accepted as an accurate or full account of our usual concept of normative reasons.

Once the importance of certain *epistemic* conditions is acknowledged, Williams suggests a more plausible interpretation of premise (2):

> **(W2)** Necessary, if A has a normative reason for ϕ then A could be motivated to ϕ when he were *deliberating* in a fully rational way departing from his existing motivational set (S). [Williams, B. 1981. p. 104, 105. 1995. p. 36[71].

By deliberation in W2 Williams understands *epistemic* deliberation partially aimed to improve our cognitive perspective on the world[72]. There are two basic advantages derived from understanding W2 in this sense.

First, W2 can explain certain intuitive cases of deliberation. Let us suppose, for instance, that the fact that a call for papers of a conference has been posted today is giving you a reason to work on an old and unfinished draft. How could be such plain non-normative fact (a call for papers) be a reason for you to act in a certain way (to revise your old draft) In the terms favoured by W2 we have a quite direct answer to this question. It is in virtue of certain psychological fact about you, that is, about your motives under certain epistemic conditions, that you ought to revise your old draft when confronted with the fact that a call for papers has been posted today. Surely, Williams could argue that the motivational importance of your own academic goals, if conveniently corrected by processes of epistemic deliberation, could be sufficient to warrant the fact that you have a reason to revise your old draft.

Second, W2 is also plausible for a more complex reason. In order to unfold this further merit of W2 let us assume, as we did in chapter 3, that our normative judgments and our motives are reliably connected. Let us suppose that:

> **(NI)** If A believes that he has a reason to ϕ, then A will be motivated to ϕ [Williams, B. 1981. p. 107 and 1995. p. 39][73]

---

[71] See Smith, M. 1994. p. 156. Parfit, D. 1997. p. 100. Dancy, J. 1999. p. 14. Hooker, B. 1987. Hooker, B. Streumer, B. 2004. p. 64.

[72] W2 is affirming that the truth-conditions of 'A has a reason to Φ' are dependents on a certain counterfactual fact: *that A would desire to Φ under certain conditions of factual enrichment*. See McNaughton, D. Rawling, P. 2004. p.111 for this reading of W2.

As we noted in chapter 3, it is very difficult to explain why NI is true, i. e. why we are strongly disposed to act in line with our self-assumed normative judgments, and why we are disposed to explain the behaviour of others by assuming that NI is true. But W2 offers at least a minimal framework to understand why it is so. It seems as if by formulating W2 we had gained a conceptual framework to explain why NI is so pervasive so to speak. So, when we include the findings offered by W2 into the commonsensical formula expressed by NI we arrive, if not to a full explanation of the motivational link, at least to a clear view of the type of content we express with an ought-based reason. And it is in virtue of this content (plus certain processes, reflexive capacities, and self-referential abilities) that we are usually motivated in line with our normative judgments. In essence, if we include W2 into the general condition fixed by NI we could obtain that

(NI)+(W2) If A believes that he would be motivated to φ if he were deliberating in a fully rational way departing from his existing motivational set (S), then A would be motivated to φ. [Williams, B. 1981. p. 110. Railton, P. 2004/1986 p. 10-1 and Smith, M. 1994. p. 177]

Further - by taking into account the sort of enrichment offered by W2 on NI – we could offer a basic (if not complete) insight to support a possible explanation of why NI is so entrenched in our practices. In the terms favoured by (NI)+(W2), it is *partially* because of the fact that when A is believing that he himself has a normative reason to X what he is believing is, in fact, that he would be motivated to X if he were deliberating in a fully rational way [McNaughton, D. Rawling, P. 2004. p. 113]

Taking into account both advantages, W2 conforms an intuitive version of premise (2). W2, however, is in need of further improvement. In the next section I will further develop W2.

---

[73] Smith, M. 1994. p. 61-62 and 148

4. An improved version of *internalism*

In what follows, I will propose that W2 could be better conceptualized if we were able to distinguish three different types of conceptual components in its formulation plus a derived account of *rationality* [Smith, M. Pettit, P. 2006].

We should understand W2 as referring to:

- A component focused on the *content*, the *dynamic,* and the proper level of *idealization* being applied to the **motivational set** that enables an agent's practical deliberation

- A component focused on the *structure* and *constraints* being applied to our **deliberation** to reach a conclusion about the desirability of a certain option

- A component focused on the sort of **capacities** involved in further *recognizing* and *responding* to a requirement framed in terms of certain conditions in line with (a) and (b)

In this section I will pay attention to these conceptual components of W2. I will postpone the derived account of *rationality* that Williams favours as a result of his compromise with W2 to the end of the chapter.

4.1 The motivational set

Williams endorses three basic points about the content, boundaries and dynamic underlying the motivational set (S) referred to in W2. First, he suggests what should be included in S if W2 were true. Secondly, he qualifies what should *not* be included in S if W2 were true. And finally, he explains how the type of deliberation exemplified by W2 determines the stability of those items that conform S.

To describe the *positive* content of S, Williams starts by assuming - against some simplified versions of humeanism - that the content of S cannot be simply equated to an agent's desires.

S, Williams defends, includes more than those non-cognitive attitudes commonly associated with desires. To put it briefly: S does not only include desires but also a more complex set of pro-attitudes (goals, plans, emotions, dispositions, evaluative reactions permeated by a representative import) that could be easily incorporated into patterns of deliberation based on processes of informational enrichment [Williams, B. 1981. p. 105].

But evidently, by assuming a *positive* view about the components of S, Williams implicitly endorses a *negative* constraint about what is not included in S. And a problem follows from it. The problem could be simply posited by asking if Williams's positive account of S is doing full justice to the importance that certain *needs* (certain experiences, certain goods, certain relationships) have in our lives. If our basic *needs* (understood as a certain range of *objectified motives* [Railton, P. 2004/1986]) are so important, why are they excluded from the items that constitute S?

The answer Williams gives to this general concern is quite straight. He simply reminds us that if an agent is not concerned with any of these valuable activities after deliberating in a fully rational way, then it cannot be held that such activities are sources of practical reasons for him. He notes:

> "If an agent really is uninterested is pursuing what he
> needs; and this is not the product of false belief; and he
> could not reach any of such motive from motives he has
> by the kind of deliberative process we have discussed;
> then I think that we do have to say that in the internal
> sense he indeed has no reason to pursue these things"
> Williams, B. 1981. p. 105

Finally, Williams notes a further point about the degree of *stability* of S as result of the deliberative processes that define W2. S cannot be understood as a static motivational set if Williams is right. Sometimes, when A deliberates about what to do – by correcting his non-evaluative beliefs and by letting his motivational states to be responsive to these epistemic improvements – certain ends that concerned him initially are not desirable once A has gained a better or a more informed perspective (for instance, knowing better the consequences of

endorsing some goal or ideal[74] [see especially Williams, B. 1981. p. 104. Smith, M. 2004/1995. p. 22].

4.2 The deliberative processes

As I understand the topic of deliberation in Williams's favoured account of *internalism*, it should be divided in two basic questions. The first one is directly concerned with the type of *principles* that could guide our deliberation. Which rules underlie the processes that enable us to discover that a certain fact gives us reasons to act in a certain way or to form a certain attitude? The second one is focused on a different problem, which relates to the type of connection established between what would be desirable to do in terms of W2 and the set of deliberative processes that determine our reasons under ideal conditions. As both questions are fairly abstract, let me introduce them in turn.

(a) The structure of deliberative processes

Williams supplements W2 by enriching its instrumental core to offer a more nuanced account of the components of our practical deliberation. Assuming the validity of the instrumental or procedural schema[75], Williams wants to explore a richer view of what deliberation is by suggesting that a wider set of processes should be added to the instrumental mould of deliberation [Williams, B. 1981. p. 104].

But which processes should be included in W2? Remember that Williams assumed that an agent partially exercises an epistemic capacity when she deliberates about what to do. This capacity is enabled and later justified by appealing to the status of her beliefs. So, in Williams's basic account, the outcome of any deliberative process carried out by A is determined, to a great extent, by certain epistemic requirements of her beliefs. Among these requirements, Williams notes the centrality of three minimal desiderata:

---

[74] And the other way around: those actions apparently undesirable under not fully rational conditions could be endorsed as desirable after a deliberative process has taken place on an agent's motivational set

[75] A *procedural* account of deliberation could be understood either as (i) an account of the *reasons* A has or (ii) as a criterion that help us to fix the degree of *rationality* of A's attitudes at a given moment. In terms of (i), A has *reasons* to Φ only if Φ could be desired after a set of deliberative process (ones departing from A's previous beliefs and desires and governed by certain epistemic standards ideally constructed - E) has taken place. In terms of (ii), A is *being irrational* in doing Φ if the deliberative processes by means of which he has arrived to his evaluative attitude toward Φ are falling short of E - even if doing Φ is desirable in terms of (i). See Hooker, B. Streumer, B. 2004. p. 58-59.

- First, the agent must have *no false beliefs.*

- Second, the agent must have *all relevant true beliefs* in her situation

- Finally, the agent must be able to *deliberate rationally* to reach to new conclusions from beliefs corrected along the previous lines [Williams, B. pp. 103-105. Smith, M. 1994. p. 156. 2004/1995. p. 22]

Let us suppose that A desired φ and A's beliefs fulfilled the previous set of constraints. In Williams's sense, A has a reason to φ because of the way desiring φ is related to certain ideally constructed psychological facts about her desires under conditions of full epistemic rationality. But there is a problem here. These epistemic desiderata are still giving voice to an instrumental or procedural account of deliberation. So, where is the improvement promised above? To answer this question I will assume hereafter that Williams widened the limits of the instrumental model by supplementing it with a set of deliberative abilities, which are not directly subsumable into the operations carried out by a simple instrumental model of deliberation. He argued, for instance, that a rational deliberator, besides of respecting the desiderata just noted, should be able to exemplify a further set of *capacities* [Williams, B. 1981. p. 104-105]:

- *First*, a rational agent must be able to temporally rank his aims and goals. He should be able to assign temporal priority to certain ends in virtue of how their satisfaction would be required in order to meet an agent's further and even more basic ends.

- *Second*, a rational agent must be capable to combine the satisfaction of several desires along with some high-level requirements such as considerations of economy, simplicity, and so on.

- *Third*, he must be able to find *constitutive* solutions, i.e. he should be capable to determine in a clear way what sort of goals to pursue in light of his unspecific or maximally general desires.

- *Finally*, a rational agent must be equipped with a certain level of *imagination*; he should be able to imagine "what it would be like if something came about" in order to further supplement the abilities previously noted. [See Hooker, B. 1987 and Smith, M. 2004/1995. especially p. 23 for a further constraint related to the importance of the *unity of a desiderative profile*].

These deliberative *capacities* are clearly far beyond of the ability to select a motivational item in virtue of how it respects certain epistemic constraints. And consequently, as far as Williams's appeal to these capacities expresses a significant addition to the core of the instrumental model, he is able to resist the criticisms that stress the simplicity of instrumentalism as the main reason to reject it as a conceptual framework. Williams argues that when instrumentalism is supplemented by the set of principles mentioned above it is free of any easy rebuttal based on its simplicity [Hooker, B. 1987].

I will return to the topic of deliberation in the next section, where I will consider some objections to the image of deliberation that Williams is offering. But as far as we are concerned now with the type of deliberative structure that Williams favours - whether it is an instrumental or a non-instrumental schema - let me establish a minimal point: Williams's account of deliberation - insofar as he focuses on some abilities that cannot be reduced to causal interactions among beliefs and desires –gives voice to a substantive departure from a simple, narrowly conceived model of instrumental deliberation.

(b) The logical form of the advice facilitated by deliberation

Michael Smith has recently developed a different approach to separate Williams's account of deliberation from simple instrumental models [Smith, M. 2004/1995 and Smith, M. Pettit, P. 2006]. Smith asks for the *logical form* of the advice reached in deliberation, highlighting a dimension that is very important to evaluate the plausibility of Williams's overall program. Asking for *the logical form* of W2 is equivalent, in Smith's sense, to ask if the normative reasons that concern W2 are a direct function of what A would desire to do *if he deliberated in a fully rational way* or, instead, if his reasons are derived from what A would desire to do *in a situation other than ideal if he deliberated from a fully rational stance*. Smith argues that the availability of these options suggests that there are two alternative ways to understand the deliberative schema proposed by W2. In respect to its deliberative import, W2 could be

understood either as referring to an *exemplar* model of deliberative advice or as pointing toward an *advisor* model of deliberative advice.

The basic difference between these alternative models goes as follows. While the *exemplar* model does *not* take into account any other motives than those being extracted from an idealized process of deliberation, the *advisor model* considers a certain amount of information about A's actual motivational profile[76] to determine accurately his reasons [Smith, M. 1994. p. 151. 2004/1995. p. 18. 2004/2001. p. 46-47. Railton, P. 2004/1986. p. 11-12]. So, while in the terms favoured by the *exemplar model* A has reasons to do whatever he desires in a context of full rationality, in the *advisor model* A only has reasons to act on those motives that could be consistently aligned with certain facts about A's *current* motivational profile.

As far as certain emotional disturbances conform a quite pervasive phenomenon in our life as agents [Smith, M. 1994. p. 134-135. 2004/1995. p. 19. See Bratman, M. 2007/2000. p. 22-23 for an overview of these undermining factors], Smith recommends the *advisor* rather than the *exemplar* model to unfold the logical form of our normative assessments in terms of reasons [Smith, M. 2004/1995. p. 18][77]

Smith defends that as far as A's *current* emotional dispositions could constrain the applicability of a certain piece of advice on him (because, let us say, the type of motive required to implement the advice is located beyond A's normal capacities of self-control and self-management), the piece of advice is normative by A's own lights only if it has been formulated by taking into account A's current emotional dispositions. Thus, although W2 offers by itself a general framework to conceptualize A's reasons at a given time, the point just added about the *logical form* underlying W2 accentuates the importance of A's current motivational profile (his operative motives, patterns of emotional activation, character-traits, and so on) to determine such reasons. It is very important to distinguish, if the remarks above are sound, between the type of advice that W2 would favour *if it was applied from an abstract perspective* and its advice *when we take into account the agent's motivational profile*. I am

---

[76] For instance, certain dispositional patterns, some facts about his previous emotional responses in similar situations, certain emotional moods that define him as agent, etc…

[77] When Smith introduces the pair *evaluated* and *evaluating* as referring to possible worlds, he points, some lines before the paragraph just quoted: "We are to imagine to possible worlds: the *evaluated* world in which we find the agent in the circumstances he faces, and the *evaluating* world in which we find the agent's fully rational self. In these terms, the internalism requirement tells us that the desirability of the agent's ϕ-ing in the evaluated world depends on whether her fully rational self in the evaluating world would desire that she ϕs *in the evaluated world*" Smith, M. 2004/1995. p. 18. (Cursive added.)

going to assume hereafter, as I noted, that W2 refers to the second type of advice [Smith, M. 2004/1995. p. 18, 25]

(c) Recognition and response-based capacities

Previously I focused on the importance of an *agent's motivational capacities* to fix the scope of W2. In the current section I am going to press a bit further on the centrality of the agent's capacities to *recognize* and *respond* to reasons. This question would require a treatment far more extensive than the one I can offer here, so I will simply sketch how W2 could confront this question, i.e. how W2 could conceptualize our psychological abilities required to recognize and respond to reasons (see chapter 8, section 1 and 2 for an additional discussion on this topic).

   On the one hand, if W2 were offering a plausible account of the type of advice expressed by a reason-based demand (R), the agent to whom R is ascribed or applied should exemplify a general capacity or set of capacities to *recognize* the *modal* fact expressed by R - concerned with his own motives under conditions of full rationality. In line with this general condition I will label the set of psychological abilities involved in recognizing a normative claim in line with (W2) as *recognitional capacities* [Bratman, M. 2007/2000. p. 24-25]. I will suggest that in order to *recognize* the modal claim expressed by W2 an agent should be able to exemplify at least two types of psychological abilities:

   (a) He should be able to *pretend* or to *imagine* a situation about how his own desires or preferences could evolve under certain processes of informational enrichment [Williams, B. 1996. p. 36. Nichols, S. 2004. p. 9-10. Railton, P. 2006b. p. 276-277 and 2006a. p. 23-24]

   (b) He should be able to gather *factual* and *personal* information to determine whether the property *tracked* by a reason-statement - whose reference is primary fixed by the set of process assumed in (a) - is being correctly exemplified. These abilities would further support the weighing processes required to pick one option among the set of possible choices supported by reasons [Smith, M. 2006. Railton, P. 2004/1986. p. 11-13. 2006b. p. 205].

On the other hand, W2 points to a *recognitional* ability (or set of abilities) essentially connected with the agent's will - at least if, as we have assumed all along this investigation, normative judgments are reliably followed by motivation. I will label this capacity or set of capacities as *responsive-oriented capacities,* and will equate them with psychological abilities that could explain how our motivational states are reliably aligned with our normative opinions. In a minimal sense, *responsive-oriented capacities* might refer to two types of psychological abilities:

(i)   They should refer to a *positive* set of agential and sub-agential psychological mechanisms (memory, attention, emotions, character traits, feelings, ideals, etc…) aimed to support and coordinate the non-deviant alignment between our normative judgments and our motives.

(ii) They should require, as a *negative* condition on agent's overall psychology, an absence of certain motivational disturbances beyond the agent's normal capacities of control – compulsive behaviour, extreme volitional disorders, extreme laziness, and so on.

However, by merely suggesting the psychological scaffolding that recognitional and response abilities could facilitate I have not answered yet a basic question: Why should we include these capacities in W2 after all? Why would we need to focus on *responsive* and *recognition-based capacities* to properly define W2 - once that we have precisely defined the content of S and the structure and logical form of our deliberation under W2? To answer to this question I will introduce a case illustrating the centrality of these abilities to grasp the way in which W2 determines our reasons in a given context. Although my favoured scenario involves only *recognition*-based capacities, I will assume that a similar case could be used to support the inclusion of *response*-based abilities in W2.

Let us assume that a certain agent (A) has a good reason to do something. He has a good reason to write an abstract for a conference to be hosted in Granada next September. In Williams's sense, to write an abstract for the conference would be a desirable thing to do for A *if A were fully rational*. To put it more precisely: to write an abstract for the conference would be advised by A's *rational counterpart* if placed in A's *current situation*. Let us

suppose that A has no stronger reasons. Besides, let us agree that A does not suffer from any motivational disturbance that could impair his behaviour in light of a self-assumed obligation. Could we affirm here that A has a reason to ϕ? By W2 we should simply claim that A has a reason to Φ, *full stop*. But let me complicate the case a bit more by enriching the description of A's situation. Let me suppose that (1) the deliberative processes carried out by the improved version of A (A+) are not extremely complex, and require simply the actualization of a set of deliberative capacities that falls entirely under the average abilities of normal agents and that (2) such average level is not instantiated, *in this particular case,* by A. Shall we say then – if (1) and (2) were the case - that A has a reason to write an abstract? [Scanlon, T. 1998. p. 369].

My intuition here is that if A cannot recognize a certain reason because he actually lacks a certain capacity (C) and C is falling under the average abilities of normal agents, then we should say that A has no reason to act the way required - at least from an internal perspective [Scanlon, T. 1998. p. 371. Shafer Landau, R. 2003 p. 177 and Smith, M. Pettit, P. 2006].

If my intuition is sound, W2 should be supplemented by including an explicit reference the possibility posited by these cases of deficiency in our normal capacities of recognition and response. W2 should be reformulated as saying instead:

(**W2'** ~minimal~) Necessary, A has a normative reason (R) supporting ϕ-ing only if W2 and (i) A *could recognize* the normative command if his deliberative and imaginative capacities were working normally and such capacities *are* working normally and (ii) A *could respond* to the normative command by forming a motivation to ϕ-ing if his motivational capacities were evolving normally and such capacities *are* evolving normally [78]

Or equivalently:

(**W2'** ~minimal~) Necessary, A has a normative reason (R) supporting ϕ only if A would be motivated to do ϕ if he deliberated in a fully rational way that departed from his existing motivational set (S) and (i) A *could recognize* R if his deliberative and imaginative capacities were normally working and such capacities *are* normally working and (ii) A *could respond* to R

---

[78] See Scanlon, T. 1998. p. 369-371. Jay Wallace, R. 2005. p. 93. Scanlon writes (p. 370): "These conclusions must be cited as showing the strength of Williams's internalism: it gives the right answer about what reasons we have because it sticks to the idea that reasons must be reachable by sound deliberative routes from agent's *actual* S (*taken to include dispositional elements*). Cursive added.

by forming a motivation to φ-ing if his motivational capacities were normally evolving and such capacities *are* normally evolving

Leaving aside some important problems regarding the meaning of *normal* in this context, I will assume that either W2' $_{minimal}$ or W2'$_{non-minimal}$ are summaries of Williams's version of premise (2). So, I will assume hereafter that any version of W2' constitutes a proper basis to discuss some points related to an internal interpretation of premise (2).
.

5. Two basic arguments in support of W2'

W2' can be defended by appealing to two different arguments [Williams, B. 1995. p. 38]:

- An argument based on the explanatory nature of normative reasons.

- An argument that appeals to our practices of blaming people.[79]

The first attempt to defend *internalism about reasons* shares a common structure with an argument than I introduced in chapter 3 (see chapter 3, sections 2 and 3) in support of *motivational internalism*. While I introduced the argument there to highlight the psychological status of our moral opinions, it could be used now to determine the grounds of our normative statements by reflecting on certain features of our explanatory practices and how they interact with an account of action-explanation widely assumed [Korsgaard, K. 1986. p. 11. Shafer Landau, R. 2003. p. 178. Williams, B. 1981. p. 106-7. Smith, M. 1987. 1994. Chapter 5]. This commonality makes unnecessary that I devote lines to introduce the explanatory argument in support of normative externalism. As far as the first argument in

---

[79] Shafer Landau, R. 2003. p. 176-181 summarizes four different arguments that support reason-internalism: (i) an argument from an action-explanation link that underlies our reasons-based statements, (ii) an argument from the dubious practice underlying a criticism of an agent's behaviour on the basis of reasons whose existence he denies, (iii) an argument stressing the close relation between blame and reason and, finally, (iv) an argument focused on the undesirable consequences that arise from analyzing certain conflicts as ones where a moral demand is opposed to an internal demand of rationality. Instead, Hooker and Streumer restrict the arguments in support of internalism to (a) an explanatory-based argument in line with (i) and (b) an argument focused on the content of our normative thoughts. Hooker, B. Streumer, B. 2004. p. 64-66

support of internalism about reasons shares a common argumentative ground with the argument for *motivational internalism,* we can imagine how we should proceed to support *internalism about reasons* by appealing to explanatory considerations (see chapter 6, section 3.1). Thus, I will be mainly concerned in what follow with the second type of argument in support of W2' - one that I will refer to as *the argument for blaming people* [Scanlon, T. 1998. p. 365].

The argument I am interested in moves around our pervasive practice of "blaming agents for specific acts or omissions" [Williams, B. 1995. p. 40]. Importantly, I am assuming that we all share an intuitive grasp of the nature of these practices of blaming agents for specific acts. And I am assuming that such grasp is partially secured because of (i) the paramount importance we assign to these practices in our daily interactions with other agents and (ii) our pervasive tendency to reflect on these practices to isolate their core features [Strawson, P.2003/1962. p. 65. Scanlon, T. 1998. Ch. 4]. Once the importance of these common intuitions is assumed, the argument starts by establishing that our practices of blaming others (by some act or omission) implicitly assume the validity of a general principle:

> **(B)** If A is blaming B for not φ-ing, then we could assume that B ought to φ-ing [Williams. B. 1995. p. 40]

B is usually connected with two additional claims. The *first one* assumes that the normativity or stringency of the *ought* included in B offers *advice* to B after his not doing φ and also guides the behaviour, sentiments, and feeling of those agents that were present when B did not do φ. Williams writes, for example: "(…) 'ought to' in the mode of blame correspond to 'ought to' in the mode of advice" [Williams, B. 1995. p. 41]. The *second* feature associated with B says that the *ought* explicitly endorsed by B is conceptually linked to reasons. Most of the times A is blaming B because he thought "the agent had a reason to act in the desirable way but failed to do so" [Williams, B. 1995. p. 41].

Once the contours of *focused blame* are assumed, Williams's argument goes as follows [see Shafer Landau, R. 2004. p. 181]:

(a) If A is blaming B for φ-ing, then A ought not φ-ing – *conceptual claim about blame.*

(b) If B ought not ϕ-ing, then B could have done otherwise than ϕ-ing.

(c) If B could have done otherwise than ϕ-ing, then such avoidance must be licensed by B's subjective motivational set.

(d) Therefore₁, if B is properly blamed by ϕ-ing, the fact that B could have done otherwise must be licensed by B's subjective motivational set.

As it stands, this argument does not support *internalism about reasons* directly. But if we add an additional premise to it, we can derive a consequence about the grounds or status of our normative reasons. The premise assumes that

(b') If A ought not ϕ-ing then A has a reason for not ϕ-ing – *conceptual claim about 'ought'*

But once (b') is added, we can interpret it as being committed with a further consequence, far more relevant for our present purposes:

(d') Therefore₂, if A is properly blamed by ϕ-ing, the fact that A has a reason for not ϕ-ing must be licensed by A's subjective motivational set.

Even though the argument above establishes the status of our reasons in a quite elegant way, it faces an important problem, as Williams acknowledges:

> "But it may seems a rather obvious fact about blame that someone can be blamed even though his S does not contain anything that would lead to the appropriate motivation; we can blame a man (we may think) for neglecting his wife even though he has no motivation to be concerned about his wife. So, if blame is necessary connected with reasons it seems to be necessary connected with external reasons" Williams, B. 1995. p. 41

When Williams is confronted with this commonsensical intuition, he reinforces his argument by moving in two different directions. The first one pursues a reinterpretation of the external intuition from an internal perspective. The second direction, a far more negative one, argues that because externalism cannot make sense of the conceptual connection established between our *blaming* A and A's *recognition* of the grounds that support our blaming him, we should reject externalism - even if our intuitions, in this context, sometimes drive us toward such position.

This negative route can be formulated indirectly by asking a simple question: If blaming an agent may be associated sometimes with a purely external sense of reasons, what is the difference between such instances of blame and simple cases of *browbeating*? In sum, how can we separate a moralistic use of blame from a more directive use that aims to correct agent's future actions? [Williams, B. 1995. p. 44 and Shafer Landau, R. 2004. p. 176]

The positive route to refuse the previous intuition requires, instead, further development to make it work. At the end of the day, however, it expresses a movement more interesting for the overall prospects of internalism. This route can be founded, for instance, on Williams's response to a different objection from externalism [Williams, B. 1995. p. 39-42]. In terms of this criticism, if our blaming others were understood as internalism defends, there would be some situations where the faulty agent would be reluctant to be blamed by invoking his incapacity to grasp the reason being adduced by us in blaming him. He could say, for instance, that no one among his actual desires, or among those that could be formed after deliberation, could be aligned with the reason that we presuppose. Williams tries to accommodate this objection into his framework. But the interesting point is that in doing so he specifies a-more-external-albeit-still-internal explanation of the scope of our blaming practices. He writes:

> "(…) blame consists of, as it were, a *proleptic* invocation
> of a reason to do or not to do a certain thing, which
> applies in virtue of a disposition to have the respect of
> other people. To blame someone in this way is, roughly,
> to tell him he had a reason to act otherwise, and in a
> direct sense this may not have been true. Yet in a way
> that it has now become true, in virtue of his having a

disposition to do things that people that he respect expect

of him (…)" Williams, B. 1995. p. 42

If Williams is right, even when our faulty agent has no deliberative-based route to grasp the reason that supports our blaming on him, the simple fact that he is blamed offers him a consideration that is intimately connected with a general disposition, an internal disposition, that any sensible agent should instantiate, "(…) to do things that people he respect expects of him" [Williams, B. 1995. p. 44 refers to a mechanism or to 'a distinctive ethical reaction']. In Williams's sense, such general disposition expresses a *basal pro-attitude* that is transparently enacted by a certain act-type (being blamed by a person that you respect) and that will guide the agent's later actions intelligibly.

So, even if there were no reasons in terms of W2' to support our blaming a certain agent (because he would not be able to grasp the reason being adduced by us either from his actual nor from his improved perspective), if the agent had reasons to suppose that those who are blaming him have a privileged epistemic stance (deserving credit for it), then we could affirm that he has an internal reason (one taking into account the basal disposition noted above) to alter his behaviour [Williams, B. 1995. p. 42-43]. Thus, even the cases that we could understand more naturally along external lines could be interpreted as supporting internalism about reasons.

6. Looking forward

In this chapter I have tried to introduce a qualified version of internalism about reasons. I have supported this meta-normative claim by appealing to two general arguments. In my sense, internalism about reasons is important because it is a fundamental component of a particular argument in support of moral non-cognitivism. My strategy will sound familiar by now. As in the motivational domain, the general argumentive structure we are concerned with now could be threatened if the qualified version of internalism were unable to provoke a widespread agreement.

Having said that, in the next chapter I will deal with some criticism of my favoured version of internalism about reasons.

Chapter 6

Exemplifying the impasse: normative reasons

1. Introduction

2. Arguments against W2'

    2.1 Arguments focused on the nature of the motivational set

    2.2 Arguments focused on the deliberative processes surrounding internalism

3. Arguments against the basic assumptions that support W2'.

    3.1 A general argument against the explanatory intuition

    3.2 A general argument focused on our practices of blaming people

4. Is externalism about reasons a less problematic option?

    4.1 Normativity, truth, and motivation

    4.2 Normativity and explanation

    4.3 Normativity and accountability

5. Assessing the *impasse* from a normative-based perspective

Chapter 6

1. Introduction.

In the last chapter I was concerned with a general argument aimed to overcome the *impasse* surrounding *Meta-ethics*. The argument defended that:

(1) Necessary, to be under a moral obligation to ϕ is to have a good reason to ϕ [*Rationalism*]

(2) Necessary, A has a good reason to ϕ only if A can be motivated to ϕ [*Reason internalism*]

(3) Necessary, if A is motivated to ϕ, then A must either desire to ϕ or desire to φ and believe that byΦ-ing she will ϕ [*Humean theory of motivation*]

Therefore – by (1) to (3) –A is under a moral obligation to ϕ only if either she desires to ϕ or she desires to φ and believes that by ϕ-ing she will φ.

I suggested that if the argument above were sound, it could help us to support a certain variant of non-cognitivism about our moral judgments or moral commitments. By departing from the *reason-referring* function linked to our moral opinions (premise 1) [Scanlon, T. 2007a], and by assuming a certain view about the status of our normative reasons (premise 2), we could derive a positive conclusion about the psychological state that is expressed by our moral commitments. We could solve the *impasse* surrounding *Meta-ethics,* in sum, by concluding that our moral opinions express a variety of desires.

But the argument above cannot offer a sound route to overcome the *impasse*, or so I argued in the previous chapter (chapter 5, section 1). And this is so because *at least one* of the premises included in the argument is disputed, i.e. at least one premise cannot be endorsed without avoiding a deep disagreement about its validity. I am referring, of course, to premise (2), which says that our normative reasons are constituted by facts about the desires that we could have in certain conditions of full rationality. In my sense, premise (2) represents the

most problematic step in the argument because it initiates, by itself, a further dispute on the status of our normative reasons. Premise (2) cannot secure the straight goal with which the argument is sometimes formulated because it gives voice to an assumption that can also be formulated along an opposite line, which could override the conclusion endorsed by the general argument (chapter 5, section 2). Once we assume this meta-methodological point, the aim of this chapter is easy to state. I will try to show how premise (2) is essentially problematic and how even an obvious alternative to it – *reason-externalism* – could also fall under serious criticisms to its plausibility.

There are two possible routes to illustrate the problematic status of premise (2). First, if we could prove that even a qualified version of premise (2) - W2'- faces serious problems, then we should agree that premise (2) does not giving voice to a well-entrenched intuition in support of any straight argument. Sections 2 and 3 will focus on this argumentative route. Second, if we could prove that even if the qualified version were not plausible we would not be in a position to endorse an alternative meta-normative view about what reasons are - reason-externalism –, then surely the prospects of any version of this argument could be severely damaged. If none of the obvious alternatives put in motion by the argument at the level of premise (2) work, then we should agree that the argument, as it stands, is not useful to find what we are looking for, i.e. a straight, intuitive, knockdown argument in support of a meta-ethical conclusion about the type of mental states that our moral opinions express. Sections 4 will deal with this negative statement.

2. Arguments against the qualified formulation of internalism - W2'

In the previous chapter I noted that a maximally qualified version of premise (2) above should establish the following conditions to warrant whether an agent has a reason:

> **(W2' non-minimal)** Necessary, A has a normative reason (R) supporting ϕ only if A would be motivated to do ϕ if he deliberated in a fully rational way that departed from his existing motivational set (S), and both (i) A *could recognize* R if his deliberative and imaginative

capacities were normally working and (ii) A *could respond* to R by forming a motivation to φ-ing if his motivational capacities were normally evolving

The condition above noted (W2' for short) is susceptible to fall under three different sorts of criticisms. W2' can be rejected because of (1) certain assumptions about the authority associated with the *motivational set* implicitly endorsed by W2', (2) because of the image of *deliberation* assumed by W2' and, finally, (3) because of the sense of *rationality* underlying W2'. I will summarize these lines of attack in what follows.

2. 1. Arguments focused on the nature of the motivational set

Williams faces an obvious problem derived of the intimate link he establishes between an agent's motives and an agent's/his reasons. In essence, Williams assumes that to determine the reasons we have at a given moment we should shift our attention to our broadly construed *motives*, i.e. to those motives that we could have if we deliberated in a fully rational way. To establish our normative situation in terms of reasons, Williams assumes that such motives always conform significant items. But why is this so? Why should we accept that our motives determine so intimately our normative status at a given moment? And why should we suppose that precisely the motives favoured by Williams are the ones that have the privilege to fix our normative status?

In my view, those who want to reject the prevalence of motives in fixing our reasons at a given time could opt for two different routes. As far as Williams implicitly assumes that our motivational set is composed of pro-attitudes or non-cognitive states with a certain direction of fit (whether they are simple desires or ideals, loyalties, plans or intentions), we could be propose a widening of the shape and content of this motivational set. We could recognize, for instance, the centrality of needs in defining our basic concerns. This widening would certainly imply a different account of deliberation and a further difference on the reasons that we could recognize and respond to after such deliberation has taken place. But in a more radical spirit - and even if we accepted Williams's non-cognitive bias toward the components of the deliberative set - we could be also disposed to reject the central role that Williams assigns to our motivational set, a role in which our motives (even if conveniently modified to respect certain commonsensical intuitions) are the proper focus of our deliberation. About this alternative, Robert Jay Wallace has recently written:

> "Thus, one might to suppose that deliberation begins from agent's existing motivations when the agent expressly takes those motivations into account, reflecting on how the ends they supply may best be attained. If we construe deliberation in this way, however, then it seems we must attach to desires the kind of normative significance that was challenges above, holding that they are themselves considerations that recommends or speak in favour of actions in the deliberative point of view"
>
> [Jay Wallace, R. 2005. 93]

The question Wallace brings to the foreground of the discussion is basically related to the purported *authority* assigned to an agent's desires in supporting normative ascriptions in terms of reasons. But this could be surely supplemented to include whatever version of our motivational set we are disposed to embrace. Why - Wallace asks - should we suppose that *any* component of our desiderative make-up offers a normatively significant item for our deliberation even when it may has gone through an epistemic process of depuration, i.e. even if it has survived to a process of *cognitive psychotherapy*? [Brandt, R. 1979. p. 11-13].

Remember that Williams rejected a *causal* model of reasoning (in which practical deliberation, although cognitively guided, is understood as a simple process of satisfactions of an agent's desires) in favour of a *heuristic* model (in which an agent's desires fix the initial scope of our deliberation without a priori constraining the range of possible concerns to be reached as result of a process of epistemic enrichment). In Williams's sense, W2' was committed with a *heuristic model*, which was one of the fundamental attractions of W2' [Williams, B. 1981. p. 102-103. 1995. p. 36. See Scanlon, T. 1998. p. 368 for the precise shape associated with the heuristic model].

But despite of Williams's attempts to delimitate his favoured model from a simple instrumental account of deliberation, many philosophers have rejected the *heuristic model* underlying W2' by the simple reason that it presupposes an internal account of the scope of our deliberation. They defend that the importance of taking our (actual) *desires* or pro-attitudes as a necessary point of departure to reach a normative conclusion about what to do that is explicitly formulated in W2', assumes a great deal of what internalism tries to establish [Parfit, D. 1997. p. 115]. In a certain way, they argue, internalism a priori denies that our

practical deliberation could lead to a state motivationally effective without previously taking into account our motives, pro-attitudes or dispositions [Jay Wallace, R. J. 2006/1990 and 2006/1999. p. 45-46. See also Scanlon, T. 1998. Appendix].

Assumed this framework, Wallace's point expresses a very common concern about instrumental accounts of rationality (no matter how developed they are or what precise role they give to the constraints on an agent's desires). In a nutshell, the problem is to know if our motivational set - even when it is cognitively improved – conforms a rational point of departure to ground our obligations in terms of reasons. Some people, as I will explain later, defend that the answer to this general question is negative. They claim that it would be possible to find some cases where a well-informed motive or set of motives could not reasonably ground an agent's reasons. So, even when the agent's motivational set (whatever we want to put into it) has been corrected to a large extent, we need a further argument to asseverate that such motives are normatively significant, which would offer an authoritative point of departure to ground our deliberation about what to do [Anscombe, E. 1963. p. 70 and 72-73. Quinn, W. 1993. p. 236].

2. 2. Arguments focused on the deliberative processes that surround internalism

In addition to the argument presented in the previous section that rejected the centrality of our motivational components to establish the basis for our normative deliberation, a further set of arguments is aimed to reject Williams's account by criticising the very conception of deliberation that W2' assumes.

a. Arguments questioning the precise focus, scope and nature of the deliberative process

Williams's account of deliberation faces three basic problems. The first has been partially suggested before, when I introduced W2', and it is concerned with the narrow *focus* that Williams assigns to our deliberative processes. The second problem is related to the supposed influence of our deliberative processes, i.e. with the expected effect of our deliberation over the *evolution and change of our initial desires*. Finally, the third problem for Williams's account is focused on the kind of *normative guidance* that it is being presupposed all along the way. Let me present them in turn.

As we noted before, William's meta-normative proposal about the status of our reasons is better understood as a subtle variety of *proceduralism. Procedural theories* of reasons and rationality assume that deliberation is a sort of epistemic-guided process, essentially concerned with the modification of a given motivational profile by means of our best-grounded beliefs. Procedural processes of deliberation aim to arrive to a well-supported normative verdict. The verdict, although not necessary free of error or superstition, at least should be able to guide consistently an agent's behaviour across time and in relation in concert with the behaviour of others agents [Williams, B. 1995. p. 36]

Evidently, if Williams's meta-normative proposal is equated with a *procedural* account of reasons, we should maybe assume that (at least) some problems that were traditionally associated with any procedural view of reasons and rationality would be transferred to Williams's account of reasons. And in a certain sense this is the case. One of the basic disagreements around *proceduralism*, for instance, has traditionally consisted in the narrow focus that it presupposes for rational deliberation. In terms of a procedural view, deliberation is basically concerned with a process of cognitive improvement that departs from our desires[80]. Against this basic claim, some philosophers defend that rational deliberation is *not exclusively* concerned with a process of epistemic correction that works on an agent's desires, at least if such process is entirely blind with respect to the content of the desires involved in the process. Some people assume, by contrary, that our deliberative processes should necessary take into account (besides our current desires) certain basic *needs* - or certain widely extended *moral* or *prudential concerns* - to determine our normative status at a given time [Williams, B. 1981. p. 105. Williams, B. 1995. p. 36-37. Parfit, D. 1997. p. 116. Hooker, B. Streumer, B. 2004. p. 70]

But as I suggested in the previous chapter, even if the case against the account of deliberation of W2' based on the importance of prudential or moral considerations is intuitively well-grounded, it receives a direct reply by Williams, which stresses the non-derivative character of his procedural account in determining our reasons:

---

[80] Williams writes: "The grounds for making this general point about fact and reasoning, as distinct from prudential and moral considerations, are quite simple: any rational deliberative agent has in his S a general interest in being factually and rationally correctly informed (…) but not a similar reason to write the requirements of prudence and morality. Somebody may say that every rational deliberator is committed to constraints of morality as much as to the requirements of truth and sound reasoning. But if this is so, then the constraints on morality are part of everybody's S, and every correct moral reason will be an internal reason. But there has to be an argument for that conclusion" Williams, B. 1995. p. 37

> "If an agent really is uninterested in pursuing what he
> needs; and this is not the product of false belief; and he
> could not reach any such motive from motives he has by
> the kind of deliberative process we have discussed; then I
> think we do have to say that in the internal sense he
> indeed has no reason to pursue these things" Williams, B.
> 1981. p. 105

Thus, Williams argues in relation with this first problem that *moral* concerns, *prudential* patterns of motivation, and the rest of the motives that could fall under the broad category of *needs,* should not be necessary taken into account to determine, by means of deliberation, the basis from which the reasons we have at a given moment will emerge.

The *second* important problem associated with the account of deliberation contained in W2' is related to the influence of our deliberative processes in systematically justifying our desires. Remember that Williams's account of deliberation accepts that any desire could be rejected (or accepted) trough the exercise of our deliberative capacities. This systematic justification could be achieved, in Williams's terms, either by considering the desire in light of our factual beliefs (which determine both, the coherence of the desire with respect to other goals of the agent and the consistency of the desire in terms of certain assumed facts) or, additionally, by putting into play a set of imaginative capacities that could make salient to the agent certain consequences derived from a possible motivational prevalence of the desire. Once both types of capacities are unfolded, Williams argues, some pre-reflective desires are going to be dropped out while others pre-reflective desires are going to be added to our motivational set. The overall consequence of this process will be that the agent will achieve a higher degree of integration or justification of his desiderative profile by the successive iteration of this dual process focused on his first-order desires [Smith, M. 1994. p. 158. McNaughton, D. Rawling, P. 2004. p. 120].

However appealing the previous image is, Michael Smith has recently pointed out that the use of factual information and imaginative resources to accept (or reject) a set of desires is not the only capacity that could shape our desiderative profile. In fact, we have a different sort of ability to normatively determine our motivational set. In terms of this capacity, we could add a desiderative component to S without supposing that this desire was directly related to any component of S – as it happens in cases of instrumental deliberation. Smith writes about that:

> "(…) the most important way in which we create new
> and destroy old underived desires when we deliberate is
> by trying to find out whether our desires are
> systematically justifiable (…) I mean that we can try to
> decide whether or not some particular underived desire
> that we have or might have is a desire to do something
> that is itself non-derivatively desirable. And we do this in
> a certain characteristic way: namely, by trying to
> integrate the object of that desire into a more coherent
> and unified desiderative profile and evaluative outlook "
> Smith, M. 1994. p. 158-159

The basic idea that Smith introduces is that as far as it is possible to conceive or to form a belief about the consequences for us of adding a new desire to our motivational profile (in terms of unity and comprehensiveness), we could add a new desire to S because of the simple fact that we believe that such desire would contribute to the unity, comprehensiveness or coherence of S. Once that we believe that it would be *desirable* to desire D in light of our current and *deliberatively improved* desires, this very belief could cause us to have a new desire for D if a certain disposition to be coherent were assumed. We would be irrational *by our own lights* if we were not forming a desire for D. Thus, while our imaginative capacities could add new desires and reject old ones by means of an simple exercise of pure epistemic abilities - " via vivid presentation of the fact" about our current desires [Smith, M. 1994. p. 161] - a process of desiderative *unification* or *integration* would add new desires trough the role played by a normative belief focused on the unity and coherence associated with our desiderative profile.

Thus, Smith defends three basic claims related with this general capacity. He assumes that (i) an agent could be disposed to form a desire for D by means of the process of *integration* depicted above, in which a normative belief plays a basic role, and that (ii) this process *cannot* be simply equated with the deliberative processes that are constrained by the availability (either direct or instrumental) of a given desire in our current motivational profile [Smith, M. 1994. p. 95 and 159]. Moreover, by recognizing the important role played by certain beliefs about the overall coherence of certain desires to conform our motivational profile, Smith assumes that (iii) a certain relativistic view could be minimized. In the terms chosen by Smith

to present this further advantage of his own favoured account of *motivational constitution*, the process of desiderative unification that he highlighted could serve to improve the prospects of an interpersonal agreement at the level of our desires, i.e. the prospects of an interpersonal agreement with respect to the reasons we have. In Smith's terms, some unjustified desires that could survive the process of deliberation depicted by Williams[81] could be dropped because they do not fit well, in terms of coherence, with our overall motivational profile. Our deliberative ability to drop or to form a certain desire in response to a self-acknowledged belief about the coherence of our desiderative profile could support this process of unification, which undermines the prospects of relativisms [Smith, M. 1994. p. 164-174]. But insofar as W2' does not take this deliberative ability into account to define the scope of our deliberation – W2' should be rejected. Or so Smith argues.

Finally, Williams's account of deliberation - as it is presupposed in W2'- faces another problem, recently stressed by Robert Jay Wallace. Suppose, Wallace argues, that we accepted a simple condition on rational deliberation, denominated by him *guidance condition*.

> **(GC)** Deliberation on our reasons must be capable of controlling *directly* our motivations and actions. [Wallace, R. J. 2006/1999. p. 46-47 and 53.]

In Wallace's terms, any theory of reasons should explain how it would be possible to be motivated to act in line with one's views about the reason one have. In sum, how "rational activity should be directly controlled by the agent's own deliberative grasp of what they have reason to do" [Wallace, R. J. 2006/1999. p. 48]. Thus GC is important in this context because Williams's account of deliberation does not satisfy it.

Wallace claims that in cases of deliberation *á la* Williams, an agent (X) is confronted with a certain fact about which he deliberates. X starts by forming a belief about such fact – for example, X believes that the bin is full of rubbish. Once he has formed this belief, X initiates a process of deliberation, which departs from his own motives and is regulated by certain epistemic constraints. If it develops normally, this process causes a desire in X to take the bin outside. I will refer to this desire as D. The important point in this rough example, Wallace argues, is that there is no stage where the agent must entertain a concrete normative thought to arrive to D. There is no direct reference to any sort of normative belief whose content we

---

[81] Because the agent is not able to exercise the best version of his imaginative capacities in order to reject them.

should take as causally effective to form D (*that the fact that the bin is full favours taking it outside* or *that taking outside the bin could be desirable*, etc…). So, although the fact that the bin is full of rubbish is normatively significant to ground A's deliberative process (because it causes X's belief), a normative thought of X (X's grasp that such particular fact has a normative status) does *not* play any role in causing a desire to take outside the rubbish. Wallace argues that we cannot find anything close to a self-determinative or self-reflexive stance from which the agent weights the different available reasons to arrive to a normative judgment, being responsive to his own reasons to self-construct his motivational profile, at any stage in this process [McNaughton, D. Rawling, P. 2004. p. 110. Wallace, R. J. 2006/1999. p. 45-51, especially at p. 50]

But how could Williams sustain his own negative view of the role assigned to normative beliefs after Wallace's suggestion has been introduced? The fact is that there are no many places to look for a clear response to this concern in Williams's work. However, and leaving aside this general problem, it seems to me that Williams suggests, at least, a general path to explain why contemplating a role more positive for normative beliefs when we deliberate about what to do could be problematic. Williams's partial response goes as follow.

First, Williams sometimes suggests [Williams, B. 1981. p. 110] that a rational agent *cannot* fully entertain a normative thought (that there is a reason for me to ϕ-ing'; 'that putting outside the bin could be desirable') except when he is deferring in someone else's authority about a certain issue. But how should we understand this apparently false claim? A plausible route is to argue that Williams is defending this controversial assertion by acknowledging that, when an agent claims to understand a normative thought, what he is doing, in fact, is to believe that some action or attitude would be desirable under some evaluative dimension. He does not believe, by contrary, that a certain fact requires or favours a certain action, or that such feature makes desirable a certain response. An agent's deliberation is a first-order deliberation, i.e. deliberation focused on the content of the attitudes [Watson, G. 2004/1975]. But if this interpretation of Williams is plausible, the point stressed by Wallace in GC should be interpreted with some caution. After all, if our deliberation is focused on the content of our attitudes and not on any essentially normative fact, why should we accept Wallace's guidance condition, one moving around the role played by an explicit normative thought? [Williams, B. 1981. 107].

But Williams tries to dismiss GC (besides of by rejecting any appeal to normative beliefs in framing our deliberation) by means of a second route. Let us suppose, Williams argues, that we assumed that normative thoughts are reliably connected with our will. Williams asks: How could a normative thought (which refers essentially to an agent's available actions) be normatively significant? How could it explain further why we act in a certain way, if it does not refer to our own motives? If Williams is right, agents are able to reason as GC requires - acting on the reasons they have - only because their motives can sometimes frame the domain of items (actions or intentions) that are normatively salient in the relevant context. If we assume that this type of self-determination is not sufficient to support GC, then Williams argues that we should explain the connection between our normative thoughts and our motives, without assuming that our motives are included in our normative thoughts at some level. To understand a normative thought framed in terms of reasons, Williams assumes, is simply a matter of acknowledging how certain items (natural ones) are related to certain properties that our motivational set exemplifies with respect to those items. They are not related to a self-conscious assessment made by the agent, as Wallace assumes.

b. McDowell's argument (or the centrality of non- deliberative capacities in deliberation)

A further criticism on the account of deliberation underlying W2' can be found in some remarks that John McDowell wrote in a paper aimed to reject Williams's conception of reasons [McDowell, J.1998a]. McDowell starts by noting that, in Williams's sense, any externalist theorist about reasons is under a clear problem related to the nature of the deliberative processes that underlies the deliberative stance that he assume [McDowell, J. 1998a. p. 99]. However, from McDowell's perspective, the real problem is not the conception of deliberation that externalists endorse, but instead the general assumptions that Williams assumed in his formulation of the supposed problem for externalism. He writes:

> "The crucial question is this: why must the external
> reason theorist envisage this transition to considering the
> matter aright as being affected by *correct deliberation*?
> He cannot make sense of the motivational effect of the
> transitions by crediting it to deliberation 'controlled' by
> prior motivations, since that would merely reveal the

reasons to be internal. So, if there must be deliberation –
reasoning – that could bring about the transition, he needs
to invent an application of reason in which it can impel
people without owing its cogency to the specific shape of
prior motivations; and this is what Williams rightly says
it is hard to believe. The argument debars the external
reason theorist from supposing that there is no way to
effect the transition except one that would not count as
being swayed by reasons: for instance, being persuaded
by moving rhetoric, and, by implication, and conversion.
But what is the ground for this exclusion?" McDowell, J.
1998a. p. 99-100

Let me note that both, Williams and McDowell, agree on one basic point: In deliberating, "the agent should acquire the motivation because he comes to believe the reason statement (…) because, in some way, he is considering the matter aright" [McDowell, J.1998a. p. 100]. Accordingly, both assume that the connection between a normative thought and a further motive is a non-deviant type of relation. It is *because* the agent entertains a normative thought - because "he is considering the matter aright" – that he is motivated to act in a certain way. But what is implied by this quite general condition? It all depends, evidently, on the account of deliberation we endorse and on how it would shape our understanding of the state being referred by our 'considering something aright'.

On the one hand, Williams defends a framework for deliberation familiar by now. In Williams's sense, at t_2 A is properly motivated to act in line with the normative thought that he had at t_1 only if A has gone trough a deliberative process - which departs from A's motivational states - and such process has been constrained by certain epistemic rules. In terms of this general account, whether A *is considering matters aright* when his motivational state is being formed or, to put in another terms, whether A is motivated in a non-deviant way by a self-acknowledged normative judgment, will simply depend on the deliberative process that he has carried out. If it has fulfilled all the conditions fixed by W2', then we can say that A is motivated to act because he has been considering matters aright.

By contrary, McDowell's approach is quite different. He basically defends that in *considering matters aright* at the moment his motivational state is formed, A does not need go

trough a deliberative process constrained by certain rules (epistemic-oriented rules) nor to take into account his previous motives to sustain such process – whatever they are. In essence, as far as A's reasons are neither ontologically nor epistemically related with his desires, it should be possible for A to have a normative reason and to act further on such reason *in a non-deviant way* even if no processes of deliberation connected his motives to act with his previous motivational set [McDowell, J. 1998a. p. 101. Pettit, P. Smith, M. 2006].

In McDowell's sense, for instance, it could be right for A (*A would be considering matters aright*) to move from not believing that he has reasons to ϕ at t_1 to believe at t_2 that he has reasons to ϕ, if it were possible to establish that A is endeavouring *a process of reflective modification* of his own assumed normative perspective through a set of *non-deliberative* capacities [Scanlon, T. 1998. p. 368 for the sense of *reflective modification* being used here and 64-72 for a wider illustration of it]

But which sort of non-deliberative capacities are we speaking of here? McDowell is quite clear. As far as the transition from not being motivated by a reason-statement to being motivated by it is neither governed by a certain kind of deliberative process nor controlled by agent's prior motives or desires, it should be assumed that we could exemplify the transition in a non-deviant way by means of *examples*, or *persuasion*, by appealing to *rhetoric*, or by simple *inspiration* or *conversion*. Thus, if this were true, we could offer an explanation of how we are motivationally responsive to the importance of certain normative features by appealing to a fully independent account of deliberation, which rejects any necessary connection between the stance we occupy when we are *considering things aright* and our motives or deliberative capacities [McDowell, J.1998a. p. 99-100]

3. Arguments against the basic assumptions that support W2'

I have criticized W2', and by extension *reason-internalism*, by rejecting its core structural components, i.e. by attacking the paramount importance that W2's gives to our motives in determining our normative stance (2.1) and by assuming that the account of deliberation that W2' favours is deeply problematic (2.2.). But besides of this piecemeal rejection of W2', reason-internalism could be denied by attacking the basic assumptions that animate any

internal account of normative reasons. Reason-internalism could be denied, in sum, by attacking *the explanatory argument* and *the argument from our practices of blaming people* (see Chapter 5. Sections 3 and 5). As far as the two assumptions are located at the conceptual core of reason-internalism, I will devote the current section to offer some arguments questioning both of them. I hope that by proceeding in this way I can show how difficult is to defend internalism even if it is understood as a general meta-normative claim about the status of our reasons.

3.1 A general argument against the explanatory intuition

The goal of the argument discussed in the present chapter and the previous one is to obtain a general conclusion about the psychological content that our moral commitments express. For this, previously we have to locate the *reason-referring role* of our moral commitments along with a theory of motivation widely assumed. The argument goes as follow:

> (1) Necessary, to be under a moral obligation to ϕ is to have a good reason to ϕ [*Rationalism*]

> (2) Necessary, A has a good reason to ϕ only if /A can be motivated to ϕ [*Reason internalism*]

> (3) Necessary, if A is motivated to ϕ, then A must either desire to ϕ or desire to φ and believe that by Φ-ing she will ϕ [*Humean theory of motivation*]

> *Therefore* – by (1) to (3) - A is under a moral obligation to ϕ only if either she desires to ϕ or she desires to φ and believes that by ϕ-ing she will φ.

I have assumed that premise (2) is critical for the prospects of this argument. To support premise (2), we usually appeal to an argument that stresses the explanatory role played by our normative statements. The argument has the following structure:

> (a) A normative reason must help to explain why someone is acting in a certain way (*Conceptual claim*)

(b)   We explain why A is acting by citing A's motivating reasons (conceptual claim about explanation)

(c)   A's *motivating reasons* are necessary referring to A's desires (*humean theory of motivation*)

(d)   Externalism about reasons claims that in those contexts where we are explaining A's action by appealing to his motives, we are not following rules of explanation that are necessary dependent on any fact about A's desires.

Therefore, if human action has to be necessary explained by appealing to A's desires (by (b) and (c)), then no external conception of reasons could make justice to this explanatory feature.

Therefore, if (a) is true then reasons-externalism is false.

We can say several things about how both arguments are connected and how they are mutually enforced. An obvious point to be noted at first sight, however, is that a good deal of the support that the second argument offers to the one going from (1) to (3) is based on an shared assumption about the nature of our *motivating reasons* (premise (b) and premise (3)). In what follows I will focus on this shared connection, and will use it as a tool to reject at least the explanatory argument in support of premise (2).

As I suggested, both premise (3) and premise (b) acknowledge the humean shape of our *motivating reasons*. More importantly, premise (3) and premise (b) can also be considered as two local instances of a more general thesis about the nature of the explanation to our actions. In terms of this wider and maybe more neutral thesis, a *motivating reason* should be always understood as a psychological state. But how plausible is the assumption shared by both arguments when we understand it as giving voice to psychologism?

In recent times, some philosophers have rejected in a straightforward manner the psychologically-biased reading of premise (b) and (3), cancelling in advance any version of it (including the humean or non-cognitive account of motivation favoured both by premise (3) and (b) [Dancy, J. 2000. Chapters 4-6]. Robert Jay Wallace, for instance, has recently offered an overview of the argument-type to be endorsed by those who want to reject any

psychological account of our motivational reasons – and by extension the particular psychological interpretation favoured by premises (3) and (b). The argument goes as follow:

i. Agents reliably act for good reasons in a non-deviant way[82] – Conceptual claim about reasons [Dancy, J. 2000. p. 103]

ii. Because of (1), we can explain an agent's actions by appealing to the normative reason the agent is under – Derived-conceptual claim about reasons

iii. Explanatory or *motivating reasons* are usually constructed as psychological states, either cognitive or non-cognitive - *Psychologism* [Dancy, J. 2000. p. 14-15, 78-79 and 85-86]

iv. *Normative reasons* are normally identified with some features of an agent's own circumstances and not with "present psychological states of " -*Externalism about normative reasons* [Dancy, J. 2000. p. 117]

v. The reason for which we are acting must be identical to those reasons justifying our actions – *Dancy's* i*dentity thesis* [Dancy, J. 2000. Chapter 5]

Therefore, if (iv) and (v) are true, we could reliably act based on our normative reasons - as premise (i) and (ii) assume - only if explanatory reasons are *not* psychological states.

But we usually act on our good reasons – by (i) and (ii)

Therefore, explanatory reasons are *not* psychological states.

Is this argument sound? I discuss this precise issue in [Gaitán, A. 2005], where I highlight some problems related to a recent attempt by Jonathan Dancy to undermine the prospects of *psychologism* as a default position in explaining human action. As far as Dancy's proposal rest heavily on a sophisticated defence of premise (v) [Dancy, J. 2000. Chapters 5 and 6] my

---

[82] Wallace writes: "(…) agents sometimes act for normative reasons, so that the normative considerations that recommend or speak in favour of their acting in a particular way (…) are the very consideration that in fact lead them so to act" Wallace, R. J. 2005. p. 89.

proposal there is to directly question premise (v) by arguing for a duality between explanatory and normative reasons. By now I have simply noted that those aiming to reject the explanatory line in support for reason-internalism should appeal to an argument similar to the one going from (i) to (v). If sound, its conclusion would deny the basic core shared by premise (3) and premise (b).

3. 2. A general argument against the intuition focused on blaming people

As I noted at the end of the previous chapter (chapter 5, section 5), there is an additional argument in support of reason-internalism. I have referred to such argument as *the argument from blaming people*. It says that

(a) If B is blaming A for ϕ-ing, then A ought not ϕ-ing - conceptual claim about *blame*

(b) If A ought not ϕ-ing then A has a reason for not ϕ-ing - conceptual claim about *ought*

(c) If A ought not ϕ-ing, then A could have done otherwise than ϕ-ing

(d) If A could have done otherwise than ϕ-ing, then such avoidance must be licensed by A's subjective motivational set

Therefore$_1$, if A is properly blamed by ϕ-ing then the fact that A could have done otherwise must be licensed by A's subjective motivational set.

Therefore$_2$, if A is properly blamed by ϕ-ing then the fact that A has a reason for not ϕ-ing must be licensed by A's subjective motivational set.

As I noted before, this argument cannot support W2' because premise (d) assumes internalism (chapter 5, section 5). For this reason, some externalists about reasons would be disposed to accept premises (a), (b) and (c) and to embrace, in turn, an alternative reading of premise (d). But which reading is this?

In externalist's hands, premise (d) can be understood in two different ways, depending on how extreme is the ban that externalism wants to impose on our ascriptions of moral

responsibility. On the one hand, it could be possible to sustain that premise (d) expresses an *incomplete* statement of the grounds required to warrant our attributions of moral responsibility. A formulation more accurate of premise (d) would be committed with the following statement:

> (**d'**) If A could have done otherwise than ϕ-ing, then such avoidance (not ϕ-ing) must be licensed by A's subjective motivational set, even when A's *current abilities* or *capacities* were *not* able to ground the internal-based requirement commanded by A's subjective motivational set under conditions of full rationality.

Evidently, **d'** is embracing a departure from a minimalist reading of W2 (W2 $_{minimal}$) in favour of a *partially external* interpretation of it (W2 $_{non-minimal}$) - see chapter 5, section 4.2. Accordingly, under **d'** an agent would be responsible not only for those desirable actions (or omissions) that fall under the scope his normal capacities but also for those desirable actions (or omissions) that, even if not falling under his *current* abilities or capacities, would be ideally *advisable for him* to do (or to omit), leaving aside his effective limitations or deliberative impairments[83].

But premise **d** could also be understood along different lines. Rather than assuming that premise **d** is expressing an incomplete way to state our intuitions about moral responsibility, some people tend to assume that we should go beyond of the psychological limits imposed by an agent's abilities in order to asses whether he was responsible for acting in a certain way To grasp the truly external sense related to our ascriptions of moral responsibility would be equivalent, they argue, to endorse something like (d''):

> (**d''**) If A could have done otherwise than ϕ-ing, then the fact that such avoidance (not ϕ-ing) is available to A is *not* dependent of A's *motives* in any plausible internalist sense, i.e. neither when *A's motives* are being corrected by a deliberative process nor when the advice offered by those

---

[83] In my sense, **d'** is *partially external* because it offers a basis to construct our duties or obligations that falls beyond our current deliberative capacities. But **d'** is not completely external either, because it observes a minimal link between our obligations and our current motives. In terms of **d',** the basis of what is required must be modelled by considering how our desires would evolve under conditions of full rationality.

deliberatively improved *motives* is further constrained by A's *current abilities and capacities*.

According to **d''** an agent would be morally responsible for an action (or omission) even if the recognition of the undesirability of such action could not be related to his own motives. But as Shafer Landau notes, if **d''** were true then *browbeating* on other's unacceptable behaviour could be licensed. As far as *browbeating* on someone - i.e. criticizing someone for failing to adhere to certain reasons that he could not be able to recognize by taking a deliberative route starting from his motivational set – is a dubious policy, we should assume that externalism cannot offer an accurate interpretation of premise (d) beyond the reinterpretation suggested by **d'** [Williams, B. 1981. p. 105-106. Williams, B. 1995. p. 39-40. Shafer Landau, R. 2004. p. 176].

Externalists could offer two different responses to this intuition around *browbeating* nevertheless First, they could support **d''** by assuming that certain *pragmatic features* associated with our self-ascriptions of reasons are compatible with the external sense championed by **d''.** They could assume, for example, that the fact that we usually do not consider our own reasons to be dependent on our own motives would license an external reading of (d). Scanlon writes, in relation to this point:

> "Insofar that we do not think that our own reasons for
> refraining from being cruel to our spouses are dependent
> of having some 'motivation' that is served by so
> refraining, we cannot regard others' reasons as being so
> dependent" Scanlon, T. 1998. p. 367

Let us suppose that we accepted that a certain phenomenological transparency operates when we are deliberating, from a first-personal standpoint, about what to do or about the reasons that we have. And let us further suppose that this transparency offers ground to sustain an external reading of **d**. Once both points are assumed, what kind of consequences could we extract from these pragmatic considerations in relation to the set of practices that enable our ascriptions of responsibility? One point should be clear. We are *not* entitled to assume that in those situations where we are blaming another agent (or witnessing the blaming of A on B) the fact that the blamed agent cannot arrive to the required action by

means of a deliberative process is *not* a significant factor to be acknowledged in assessing the accuracy of our blaming. But at the same time, and however central the motivational link stressed by internalism is, I believe that in order to secure our blaming others we must pay attention to contextual factors wider than those focused on the availability of a route to connect the action required by our blaming and the motives of the agents being blamed. The objective *wrongness* of an action, the social consequences derived from our blaming others in those circumstances, and our emotional reactions to certain acts, must be taken into account to evaluate whether our blaming others is warranted even if they cannot do otherwise. But surely, and insofar as some of these factors can be constructed by attending to certain values, our blaming others can be warranted by these values rather than by the mere availability of a link between an agent's motives and the required action.

The second movement an externalist would propose to defend the plausibility of **d''** would stress an intuition about the kind of responsibility we have for our own motives. This route might start by defending that, other things being equal, an agent can be responsible not only for those actions that are connected to certain *improved* motives that he endorsed after a deliberative process focused on his *current* motives, but also for the precise shape and content of his *current* motives. If it was possible to prove that sometimes an agent is responsible for his own will – understanding *will* as the agent's current set of motivational states [Frankfurt, H. 1971. Bratman, M. 1999. Introduction] - then there would be a plausible basis for blaming him even if the particular action or omission required cannot be connected to any improved motive of him. In a certain way, the adequacy of blame in a given situation would be dependent, as before, on the availability of a counterfactual fact. But now the fact would be focused on the degree of control an agent has on his own motives and desires rather than on the connection that could be established between his *improved* motives after deliberation and the required action or omission.

If any of the routes in support of (d'') were accepted, we could be in position to offer a different kind of argument from our practices of blaming people, one defending externalism:

(a) If A is blaming B for φ-ing, then A ought not φ-ing - Conceptual claim about *blame*

(b) If A ought not φ-ing then A has a reason for not φ-ing – Conceptual claim about *ought*

(c) If B ought not ϕ-ing, then B could have done otherwise than ϕ-ing

(d'') If A could have done otherwise than ϕ-ing, then the fact that such avoidance (not ϕ-ing) is available to A is *not* dependent of A's *motives* in any plausible internalist sense, i.e. neither when *A's motives* are being corrected by a deliberative process nor when the advice offered by those deliberatively improved *motives* is further constrained by A's *current abilities and capacities*.

Therefore, that A can be properly blamed by ϕ-ing is entirely independent from A's motives in any plausible sense.

But if this argument is sound, then the second route to defend reason-internalism is severely debunked.

4. Is externalism about reasons a less problematic option?

So far I have characterized reasons-internalism along a double dimension. On the one hand, I have introduced a qualified version of reasons-internalism, which encapsulates a set of necessary conditions to warrant normative ascriptions in terms of reasons. On the other hand, I have presented two general intuitions that could show the attractiveness of reasons-internalism as a meta-normative theory. I have assumed that the explanatory import linked to reasons and a general intuition about our practice of blaming people could help to explain why reasons-internalism is sometimes taken to be the default option in theory of normativity. But I have done another thing. By performing the negative task just referred I have implicitly sketched the contours of what an external theory of reasons could be. In a certain way, by criticising reasons-internalism I have offered an *indirect* formulation of *externalism about reasons*, which is committed to three basic claims:

- *A basic claim about the warrant required for a normative ascriptions couched in terms of reasons*: It would be possible for A to have a normative reason to ϕ even when A would *not* be motivated to ϕ in any of the senses defended by internalism – i.e. neither

when *A's motives* are being corrected by a deliberative process nor when the advice offered by those deliberatively improved *motives* is further constrained by A's *current abilities and capacities.*

- *A related coda constraining our explanatory practices*: A's having a normative reason at t_a for ϕ is relevant to explain A's further action at t_2 even if doing ϕ at t_2 is not necessary connected with A's desires or motives.

- *A related coda coming from our practice of blaming people*: If *ought* is conceptually connected with 'could have done otherwise' in those contexts where an agent (A) is blaming another (B) by an action or omission, we may assume that the *relevant* range of possibilities associated with ought is not necessary constrained by B's motivational capacities.

But evidently, *externalism about reasons* is not only the negative counterpart of reasons-internalism. *Externalism* expresses by itself a peculiar standpoint to think about the nature of normativity, which leaves aside the intuitions and assumptions that were assumed by internalism about reasons[84]. But if externalism is by itself a positive view, why cannot we use it in order to offer a sound interpretation of the general argument that was sketched at the beginning of the previous chapter (Section 1, 2, and 6). Why, in sum, not use externalism to re-start *Meta-ethics*?

In what follows I will outline a positive version of externalism, arguing that even if we managed to construe such version it would not give us a good meta-normative framework to interpret the second type of assumption contained in the general argument described at the beginning of chapter 5 (see section 1 above). As far as the positive version is not free of ambiguities and problems, we cannot appeal to it to secure the general argument – at least if our interest is to establish a knockdown case in favour of a certain meta-ethical view. To illustrate this general point, I will be concerned with three basic types of statements any normative externalists should endorse: First, I will focus on a general thesis defended by

---

[84] For example, anti-causalists theories of action face, to some degree, the same sort of methodological problems. In the context of a causal orthodoxy widely accepted, some anti-causalists theories of action (volitional theories of human action, for example) are taken as obscures or as simply proposing capacities beyond the simple causal processes that work in a moderate naturalist framework. Evidently, the point being made here is that, to some extent, such theories should be positively concerned with much more than criticizing a causal approach to human action if they want to be taken as substantive alternatives to the causal orthodoxy.

externalism about the sense in which certain normative statements referring to reasons can be *true*. Second, I will be concerned with the way in which externalists relate *explanation* and *reason* in certain contexts, i.e. ones where we want to make sense of the behaviour of an agent. And finally, I will point some things about the limits that externalism defends on the topic related to the sort of *accountability* that our practices of blaming people express.

4. 1 Normativity, truth, and motivation

Let us suppose that an agent made the following assertion when faced with a situation quite familiar:

    (R) To help him is what I have most reason to do

I will refer to these assertions as R-type assertions. The externalist theorist is basically interested in two related questions about R-type assertions. First, he is concerned with a *semantic question*; he wants to know if the propositional content expressed by R-type assertions can be true – and if such content is true in fact. Second, he is concerned with a *psychological question*; he wants to know which psychological state we are expressing when we utter R and why this psychological state – whatever it is - is able to influence the agent's further motives. He wants to know, in sum, how the normativity associated with R-type assertions can be explained by appeal to the content of those psychological states underlying R-type assertions.

Reasons-internalism acknowledged also the importance of both concerns. Internalists argued, for instance, that the truth-conditions of a R-type assertion could be determined by a psychological fact that refers to the motives that an agent could endorse under conditions of full rationality. Let us suppose that the proposition expressed by our utterance of (R) can be paraphrased in the following way:

    (R$_p$) that I have reasons to help him[85]

---

[85] Scanlon, T. 1998. p. 57

Internalism defends that R_P is true if and only if a non-normative proposition - (RI) - is true. So, when I utter a sentence whose content is couched by R, such content is true if and only if the following proposition is the case

(R_{NN}) that I will be motivated to help him if I deliberated in a fully rational way from my current motives[86]

But internalism also connects this semantic insight with a second interest focused on the psychological nature of our normative opinions. In doing so, internalism assumes that the *truth-conditions* of a normative assertion R directly determine the *content* of the thought being expressed by R. They accept, in a nutshell, that the *psychological* content expressed by R is determined by those rules that govern truth-preserving assertions of R. Therefore, if (i) the normative proposition contained in R can be true if and only if the fact referred by R_{NN} is the case and (ii) the psychological content expressed by R is determined by R's truth-conditions, then it follows that the psychological content being expressed by R is equivalent to the following proposition:

(R_T'') [that I will be motivated to help him if I deliberated if a fully rational way from my motives]

However, if we want to offer a complete account of the psychological state being expressed by a normative assertion – one, remember, that can explain the motivational import of this type of assertions – we should be able to highlight what sort of *attitude* takes as its object the just noted content[87]. Internalism could secure the motivational link either by proposing beliefs or some kind of pro-attitude (desires, prescriptions, commitments, etc…). Some internalists assume, for instance, that in asserting R an agent is expressing a belief with the content of (R_T''). The resultant composite - *attitude* plus *proposition* - should help us to explain (with the help of certain dispositions to align, in coherent patterns, our beliefs and desires[88]) why we

---

[86] Compare with Velleman: "A reason for acting is a proposition whose truth would (…) justify an action". Velleman, J. D. 2000/1992. p. 100. See also Smith, M. 2004b. at page 78.

[87] Dancy writes: "(…) what is believed can contribute to the explanation of the action, since the believing comes with a content, and needs to do so if it is to explain anything. So, it's the psychological state plus content that together constitute the motivating reason (…)" Dancy, J. 2000. p. 113.

[88] At this stage some internalists make an important assumption - Smith, M. 1994. 2004/1995. 2004b and 2005. Let us suppose, they argue, that we were equipped with a psychological sort of capacity to align, in a coherent

tend to be motivated to act in line with our self-focused normative assertions. I will refer to this option as *the default option for internalism*. *The default option* assumes that:

- (i) The proposition expressed by (R) is true – *that I have reasons to help this needy person* - only if a certain non-normative fact - *that I would be motivated to help him if I deliberated in a fully rational way from my current motive* - is the case

- (ii) Once the agent acknowledges (R_{NN}) - forming a belief focused on such content - he would be irrational *by his own lights* if he were not motivated to help her, i.e. if he did not form a motivational state oriented to act in the way depicted by (R).

Until now I have stated the basic shape of internalism attending to its semantic and psychological commitments. How does externalism conceptualize both issues? How does externalism understand the truth and force of our normative utterances?

Let me start by the semantic question. The most basic and simple statement endorsed by any variety of externalism about reasons claims that a normative demand to act in a certain way could be truly ascribed to a certain agent even when such ascription is not tracking neither the agent's current desires nor the desires that the agent could endorse after deliberating in a fully rational way [Williams, B. 1981. p. 101-102. Parfit, D. 1997. p. 113-114. Shafer-Landau, R. 2004. p. 206]. Parfit notes:

> "On their view [externalism], we all have reasons to
> protect our health, and to prevent our own future
> suffering, and these reasons do not depend on whether,

---

way, our desires in relation to (or in respect with) those beliefs underlying the assertion of a normative claim – our believing that certain action would be desirable under conditions of full rationality. Let us refer to such capacity as *a disposition to be coherent*. Such capacity or disposition for coherence, internalists argue, would regulate our motivational responsiveness without being necessary developed from a self-conscious standpoint - in the same way that those belief-like regularities expressed by a normative assertion are not *necessarily* identified with any conscious state whose content is in line with (R_T''). If such scenario were plausible, an agent could be motivated to act in line of what is required of him by a normative assertion even if he were lacking the sort of awareness pointed before - i.e. even if a belief-state whose content is (R_t'') were absent from his own self-acknowledged deliberative perspective.

after informed and rational deliberation, we would care

about these things" Parfit, D. 1997. p 101.

In terms of externalism's basic semantic assumption, for any normative assertion to be true, it is not required to assume that a certain psychological fact about the agent might be the case [Parfit, D. 1997. p. 123. Parfit, D. 2006. p. 335-336]. Insofar as it is a quite general statement, let me consider again (R)

(R) To help him is what I have most reason to do

We assumed that the propositional content associated with (R) is

(R$_p$) that I have reasons to help her

Taking for granted the previous background, externalism assumes that (R$_p$) is true if and only if the following condition is the case:

(RE) that there are certain features that are favouring my helping her- whatever my existing motivations are

Remember that the *second movement* I made before to explain the relation between normative truth and psychological and motivational issues identified the *content* being expressed by a normative assertion (R) in terms of the truth-conditions of R. After that, we identified the psychological attitude that takes such content as object by simply assuming an explanatory condition: if the attitude proposed can respect the multiplicity of roles underlying our normative thought (paying special attention to the motivational ones) then the attitude would be a plausible choice to unfold the type of psychological state being expressed by our normative utterances. So let me focus on truth-conditions of R and let me try to specify RE a bit further.

Some people start by assuming a normative-based account to identify the truth-conditions of R. They assume that R is true only if a particular *normative fact* is the case, i.e. *the fact that certain features favour a certain attitude of an agent in a certain context*. This normative fact,

they defend, is the truth-maker of R[89]. In consonance with this intuition they offer a quite straight interpretation of RE:

> **(RE')** that some features associated with the agent's situation are *favouring* a certain attitude or action by the agent, by themselves and *whatever* existing motivations in agent's side[90] - Scanlon, T. 1998. p. 57. Parfit, D. 1997. p. 109, 111, 123, 125-127. 2001. p. 18. 2006. p. 331, 335-336. McNaughton, D. Rawling, P. 2003. p. 37.

But other people defend that R is true only if certain evaluative features can be connected with the action being favoured by a reason[91]. They defend that a normative assertion, i.e. a reason-referring normative assertion, is true only if an *evaluative fact* is the case. This intuition can be tentatively unfolded by presenting the truth-conditions of R in the following way:

> **(RE'')** that some *evaluative* features associated with the situation in which this person is located are favouring a certain attitude or action in the agent, by themselves and *whatever* the agent's existing motivations - Raz, J. 2001a. p. 22-24 and 2001b. p. 65

Finally, other philosophers tend to assume that, even if it is true that we sometimes refer to evaluative concepts to unveil our reasons, normative assertions are not true in virtue of the evaluative facts referred by such concepts. A normative assertion, they defend, is true simply

---

[89] Parfit notes: "for these statements [normative statements] to be normative they must be capable of being *true in a strong sense*" See Parfit, D. 2006. p. 332. Or consider this quote from McNaughton and Rawling: "Suppose that *P* or *Q* and that *not P*. Because of this conjunction you have a reason to conclude that *Q*, the conjunction being your reason (as distinct from the fact that you have a reason). Or suppose Eve has a headache, and Al has an aspirin that will relieve it. Because of this circumstance, Al has a reason to give Eve an aspirin: the circumstance is this reason (as distinct from the fact that Al has this reason)" McNaughton, D. Rawling, P. 2003. p. 43.

[90] McNaughton and Rawling argue, for example, that "(…) reasons might be descriptive facts (although we contend the fact that they are reasons is not descriptive) (…) We maintain (along with many others, such as Derek Parfit 1997) that there are irreducible normative facts" McNaughton, D. Rawling, P. 2003. p. 25. See also at page 34 and 37 and, more recently see McNaughton, D. Rawling, P. 2004. p. 123

[91] Compare with Raz, J. 1999. p. 90 and 106. Those supporting this account assume the necessity of making a simple clarification to accentuate the externalist credentials of these evaluative-based proposals. They assume that the set of *good-making*, *evaluative* or *worthwhile-giving features* (the *pleasantness* derived from taking a day off, the *enjoyment* derived from renting a classic movie, the *braveness* associated to a certain response to an offence made by a rude person, the *kindness* showed by a certain behaviour, the *justice* of acting in favour of a certain institutional arrangement) although importantly conditioned by the agent's goals in some occasions (*pleasantness*) need not to be exclusively understood as essentially determined by such contingent facts. In fact, some evaluative features of the set of examples just given are instantiated by some actions that could not be subjectively dependent on our concerns or desires in certain situations (the case of *justice*, *braveness*, and *kindness*) but that even in these cases could make the action a worthwhile option to be chosen.

in virtue of a descriptive fact, in virtue of some evidence. Thus, the truth-conditions governing a normative assertion are not correctly expressed neither by endorsing (RE') nor by assuming (RE''), but rather by simply referring to the following general condition:

> **(RE''')** that some *descriptive* features associated with the situation in which this person is located favouring certain agent's attitude or action in the agent, *by themselves and whatever the agent's existing motivations* - Dancy, J. 2000. p. 2-4, 99, 101-104. Parfit, D. 1997. p. 117.

Parfit writes, for instance:

> "On value-based theories (…) reasons are provided by various natural features of the object of our desires, and it is from these reasons that all other reasons derive their force (…) when an object has such reason-giving features we can call it good, but that is merely an abbreviation: a way to imply that it has such features" Parfit, D. 2001. p. 20

Once we are aware of these different paths to unfold the truth-conditions of our normative assertions, two things should be clear. On the one hand, in the discussion about normative truth a great deal of disagreement is focused on the nature of the facts that make our normative claims true (i.e. on the nature or proper characterization of normative truth-makers) rather than on the general theory of truth to be endorsed. Most of the theorists that support externalism about reasons are happy to defend a substantive theory of truth, leaving unexplored another proposals – see Chapter 2, section 3. On the other hand, and leaving aside the precise resolution of the disagreement about the nature of truth in the normative realm, a certain image is being implicitly accepted by externalists about reasons. As Jonathan Dancy has noted:

> "Those that suppose that reasons must rest on values if they do not rest on desires are implicitly imposing a certain layer-cake structure to the normative. At the bottom there are the features that generate value; above that there is the value so generated; and above that there

> are the reasons and requirements that are laid on us by the
>
> prospect of value" Dancy, J. 2000. p. 29

Having assumed both points, what can externalists say about the nature of our normative thoughts? What kind of content we entertain when we utter a normative assertion? If the path that connected the truth-conditions of an assertion and its psychological content were followed, we should start by assuming at least three parallels views about the content of the mental states expressed by our normative assertions. Additionally, we should propose a concrete attitude to fill the functional role that is usually presupposed to our normative thoughts. Let me start by the first task, the one asking for the *content* of our normative thoughts. Some people may assume that when we assert R we are entertaining a *normatively framed content*. Scanlon notes:

> "How, then, should we interpret a judgment that X is a
>
> reason for doing A? One possibility is to say that a person
>
> who accept such a judgment takes a certain belief to be
>
> warranted, namely, the belief that the relation 'counting
>
> in favour of' holds between X and doing A" Scanlon, T.
>
> 1998. p. 57[92]

And in the same vein Velleman points that:

> "When an agent acts for a reason (…) The agent's
>
> attitudes are thus conceived as having propositional
>
> objects that intrinsically favour a particular action, and
>
> their favouring the action is conceived as crucial to their
>
> behavioural influence" Velleman, J. D. 2000/1992. p.
>
> 101.

Other philosophers, on the contrary, defend that, insofar as the truth-conditions of R are fixed by RE'', when we make a normative assertion we should refer to certain evaluative features that are present in the current situation (the *kindness* of a certain reaction, the *justice*

---

[92] Parfit points, for instance: "When I believe that I have most reason to jump, I am believing that I should to jump, and that, if I don't, I would act irrationally, or making a terrible mistake" Parfit, D. 1997. p. 123

of a certain arrangement, the *cruelty* of a certain remark, and so on). If they are right, most of the times our entertaining a normative thought is enabled by our disposition to focus on certain evaluative features when we are thinking about an action, an intention, or an emotion rather than by an explicit understanding of a normative demand. Let us consider, as an illustration, this example taken from Ulrike Heuer:

> "It Saturday's night, and Sally is trying to decide whether or not to go to a party at Smiths'. The party promises to be enjoyable and would allow Sally to meet some friends (…). Evidently, she has reasons to attend to the party, and her reasons can be phrased in terms of the perceived evaluative aspects of going" Heuer, U. 2004. p. 130[93].

Finally, sometimes it is defended, in line with RE''', that in entertaining a normative thought we are focused on certain non-evaluative features, i.e. on certain facts that favour a certain action in a more or less direct way. In terms of this proposal, our normative thought is neither focused on a normative fact about how certain descriptive facts favour an attitude, nor focused on some evaluative features associated with the attitude itself. Rather, when we entertain a normative thought we are tracking a simple fact about our own situation, a fact that, if true, could give us a reason to act in a certain way [Parfit, D. 2001. p. 25 and 29. Setiya, K. 2007. p. 9[94]]

Now, let me say something about the second task that externalists should attain to offer a complete account of the nature of our normative thought. Remember that this second task consisted in stating what type of psychological *attitude* takes the array of possible contents just referred as its object. Externalists should ask now: Does the assertion that there is a reason to do a certain thing (when *reason* is considered along an externalist interpretation) express a cognitive or a non-cognitive sort of attitude toward a certain type of content?

---

[93]Joseph Raz suggests that: "Reason is then explained in part by invoking value: valuable aspects of the world constitute reasons. This approach, the classical approach, it may be called, can be characterized as holding that the central type of human action is intentional action; that intentional action is action for a reason; and that reasons are facts in virtue of which those actions are good in some respect and to some degree" Raz, J. 1999. p. 22-23

[94] I envisaged this possibility during a talk by John Broome abound the different ways to understand the expression 'responding to reasons'. See Broome, J. (forthcoming) In Broome's own interpretation, Parfit could be understood as defending a non-evaluative reading of the content associated with our normative thoughts. So, *in responding to reasons*, Parfit would be taken to defend that we are, in fact, responding to beliefs whose content is entirely descriptive or non-normative. See Parfit, D. 2001. p. 25.

Evidently, as realists about reasons, the majority of externalists assume that when an agent makes a normative assertion he is saying, in essence, how the world is [Velleman, J. D. 2000/1992. p. 101. Shafer-Landau, R. 2004. p. 206. Parfit, D. 2006. p. 332]. So, at least at a first sight, externalists are disposed to assume that, when we assert a normative statement, we are expressing a belief-like attitude [Nagel, T. 1970. p. 3]. Along these lines, Scanlon acknowledges:

> "The things that are reasons are, as I have said, the same
> kinds of things that can be the contents of beliefs –
> propositions, one might say (…)" Scanlon, T. 1998. p. 57

But evidently, by simply assuming a belief-like attitude as the proper psychological state to be expressed by a normative statement, externalists about reasons are not securing their position. Additionally, they should offer positive arguments against those that defend that non-cognitivism about reasons should be considered as the default option to unveil the kind of attitude expressed by a normative assertion. *Non-cognitivism* (a theory that I will take here as defending, until further clarification, that normative assertions are aimed to *express* non-cognitive states, e.g. states of acceptance of a certain system of norms [Gibbard, A. 1990. p. 71-80]) poses three basic problems to any cognitivist view. It is sometimes assumed by non-cognitivism, for instance, that an account not cognitively-based could better explain the normative force of our reasons-referring assertions [Blackburn, S. 1998. p. 68-69. Gibbard, A. 1990. Chapter 4 and 2003. p. 29, 152-153].

Additionally, non-cognitivists argue that some cognitive-based accounts are committed with dubious metaphysical entities, i.e. with non-natural facts of the type I previously referred as? appealing to the favouring relation [Gibbard, A. 2003. p. 16. Shafer-Landau, R. 2004. p. 18. Blackburn, S. 1998. p. 84-85].

And finally, and as result of the previous commitment, non-cognitivists sometimes adduce that certain varieties of cognitivism about reasons are unable to offer a plausible account of the processes by means of which we attain normative knowledge [Shafer-Landau, R. 2004. p. 22. Scanlon, T. 1998. p. 59].

How strong these criticisms are? And more importantly: To what extent is externalism about reasons affected by them? As I see it, externalism should be able to face at least two of

the previous criticisms[95]. As far as the content of our normative thoughts is explained by appealing to plain facts - RE''' - or by considering the centrality of certain evaluative-loaded patterns of perception - RE'' - we could cancel the metaphysical and the epistemological worries adduced by non-cognitivism. But what about the first criticism adduced by non-cognitivism? What about the criticism saying that externalism about reasons, if construed as a cognitive-based theory, is unable to explain the normative or action-guiding dimension of our reasons-referring assertions? To put it in other words: How could we render the sort of *prescriptivity* underlying our reasons-referring assertions intelligible by appealing to a cognitive sort of state? I suspect that externalism needs to work a bit harder to answer this general concern. In the next section I will try to explain how externalism explains the normative import of our reasons-referring assertions [Scanlon, T. 1998. p. 60. Parfit, D. 2006. p. 338. McNaughton, D. 1988. p. 48]

4. 2. Normativity and explanation

Externalists can offer two alternative stories to explain the guidance offered by our normative claims, i.e. their *prescriptivity*. They can explain the *prescriptivity* or *normativity*[96] of our reasons-referring assertions by means of an internalist model or they can recur to an externalist account of motivation to unfold the motivational import of our judgments of reasons. Importantly, both routes can explain the motivational import of our normative assertions without rejecting the semantic and the psychological story just sketched.

Some theorists among those recurring to an externalist schema believe that by reflecting on some functional features of an important subset of motivating-encompassing attitudes (*intentions*), we could unfold the prescriptive character of our reasons-referring assertions [Bratman, M. 1987. Chapter 1. 2001. Scanlon, T. 1998. p. 45-46]. They start by assuming that intentions are *judgment-sensitive attitudes*, i.e. attitudes that are intrinsically responsive to our judgments about the reasons that we have in a certain context [Scanlon, T. 1998. p. 21][97]. Scanlon writes:

---

[95] See Shafer-Landau, R. 2004. p. 22-37 for a wider and more detailed account.

[96] See Copp, D. 2007. Introduction for some general remarks on *prescriptivity* and *normativity*.

[97] Although judgment-sensitive attitudes conform an assorted set of attitudes (including, among others attitudes, desires, emotions, certain feelings, ideals, and other evaluative attitudes [Scanlon, T. 1998. p. 20]) Scanlon's concern about intentions could be understood, evidently, in terms of the centrality of these states of action-guiding commitment to directly refute the general intuition stating that, as a matter of conceptual necessity,

"The class of attitudes for which reasons in the sense I have in mind can sensibly be asked for or offered can be characterized, with apparent but I think innocent circularity, as the class of 'judgment-sensitive attitudes'. These are the attitudes that an ideally rational person would come to have whenever that person judged there to be sufficient reason for them and that would (…) 'extinguish' when that person judged them not be supported by reasons" Scanlon, T. 1998. p. 20

Once our intentions are located into the set of judgment-sensitive attitudes, Scanlon thinks that we could obtain a general framework to explain the prescriptivity of our reasons-referring assertions only if we accepted two further assumptions. The first one notes that by doing ϕ we are usually expressing an intention aimed to ϕ, so that "is the connexion to judgment sensitive attitudes what makes events actions" [Scanlon, T. 1998. p. 21]. The second assumption is about the focus of our reasons. It simply defends that a reason supporting our *doing* ϕ is (other things being equal) a good reason to *intend* to ϕ. Once both assumptions are assumed, Scanlon is able to offer a direct explanation of the *prescriptivity* of our normative assertions. A reasons-referring assertion is able to directly influence our action because the intentions that underlie our actions are constitutively responsive to our judgments assessing the reasons that we have [Shafer-Landau, R. 2004. p. 204]. As an obvious consequence of this general capacity associated with our intentions (*reasons-responsiveness*), we should consider the set of normative assessments endorsed by the agent as prime components to sustain our explanations of his actions [Scanlon, T. 1998. p. 21-22].

Finally, those theorist who recur to motivational externalism to explain the prescriptivity of our normative assessments, sometimes refer to a well-known story. I paid attention to this matters in chapter 4 (section 4.2), so I am not going to extend very much here. In terms of this by now well-known story, an assessment framed in terms of reasons issues the formation of an intention to act in a certain way only because a certain desire intervenes. The content of this desire compels either to do 'whatever is taken to be supported by reasons because it is so

normative reasons must include some reference to the agent's desires in order to be explanatory relevant for an action.

supported' or to do 'each option that is supported by reasons because of these very reasons' [Dreier, J. 2006/2000. p. 548]. Thus, it is because we have a certain desire that certain normative beliefs can guide our action.

4. 3. Normativity and accountability

Although it is importantly focused on the nature of normative truth and normative motivation, the polemic between internalists and externalists about reasons goes far beyond the previously noted issues. It is also concerned with some topics related to the status of the practices by means of which we evaluate each other's behaviour in terms of their *rationality* or *irrationality*.

In terms of this broader interest, internalists defend that when we evaluate the rationality of an agent (A) we are trying to make explicit the contours of an *advice* to A, which could be grasped by A in his current situation. They assume that the scope of the advice that we could offer in these cases does not go beyond the reach of the motives that the agent could form under certain counterfactual conditions [Smith, M. 1994. p. 171 and 2004/1995 and 2004/2002. p. 349].

Externalists, in turn, assume that our evaluations of rationality fall far beyond any advice intended from an *if-I-were-you-perspective*. Ascriptions of rationality, they tend to argue, do not try to shift the agent's current motives in virtue of some normatively significant fact about his own improved motives. On the contrary, they make the features that could justify a certain action salient even when they fall far beyond A's abilities and concerns [Scanlon, T. 1998. p. 372].

Evidently, once the shape of both meta-normative parties is minimally delineated, how could we decide about the disagreement just noted around the scope of *rationality?* How should we understand the limits and constraints to the evaluations of an agent's choices, processes and abilities in terms of rationality or irrationality?

Take internalism again. From this perspective, whether A is irrational in doing ϕ simply depends, *first*, on whether Φ would be desirable after a deliberative process – one along with W2' - has been carried out and, *second*, on whether A's current capacities for deliberation fall within the normal level of achievement or excellence by means of which the desirability of ϕ is determined. If both conditions are satisfied at t_1 by A and he is not being responsive to the

desirability associated with Φ, internalism should affirm that A is being irrational – as far as he is not motivated to act in line with his own better judgment. Conversely, if one of the previous conditions were not satisfied at t_1 by A, then we would *not* be entitled to assume that A is being irrational for not doing ϕ - even if we are tempted to defend that, from our own perspective, doing ϕ could be a very desirable choice [Scanlon, T. 1998. p. 370-373][98].

Externalists, however, reject the previous picture. They do not accept that certain abilities or *capacities* of A can provide a relevant informational basis to determine the rationality of A's choices in a given context. Although a certain degree of insight into A's deliberative capacities and abilities is not entirely irrelevant to assess A's merit for a certain action (or to modulate our reactions toward A's rational failures), this information is not sufficient by itself to support a genuine assessment of A's rationality. An additional amount of information posited in terms of *values* and fully independent of A's motives, abilities, and capacities – is always required to properly judge A's rationality. So, even if we accepted that A's impaired motives or emotions supply information to evaluate the rationality exemplified by A in treating his wife badly, the fact that A is an insensitive person is not sufficient by itself to cancel neither the fact that A had a reason to treat better his wife nor the fact that, if he was rational, he should had treated her better. As Scanlon notes "his inability to see this makes no difference" neither in whether he has a reason to treat her well nor in whether he is irrational by not treating her well [Scanlon, T. 1998. p. 371-372. See Parfit, D. 1997. p.109-110 and 2006. p. 336].

Even if this is plausible, a question still remains if we accepted that externalism is pressing a central aspect of our practices of appraisal in terms of reasons. We could ask the following question to externalists: If by appraising A in terms of rationality we are neither offering *advice* to A (by taking a better-placed perspective on his own motives) nor *evaluating* him in virtue of how his motives and abilities fall short from certain standards, then what is the point of saying that someone would behave irrationally by doing a certain thing? How can externalists offer an account of the function of our judgments of rationality?

As I see it, externalists could endorse an alternative function for our assessments of rationality by shifting our attention toward the *propedeutic* or *critical* character of our

---

[98] Internalism will further explain a common fact of linguistic use in those contexts. If we insist to describe the normative status of an agent's performance as 'irrational' even when one of the conditions is not satisfied, then this use of irrational we would exemplify an *evaluative* use or, in Parfit's words, a classificatory use in line with forms like 'it could be *good* or *desirable* if the agent were V - where 'V' stands for an evaluative term as 'kind', ''sensible', 'just', 'temperate', etc. See Williams, B. 1995. p. 39 and Parfit, D. 1997. p. 111.

evaluations of rationality. They could simply defend that by saying that A is irrational we are not merely advising him in a conditional sort of way - one referring to his motives and abilities – but we are rather confronting him with some facts (i.e. with some *values*) that could sustain a motivational shift in him [McDowell, J.1998a]. When we say that A is irrational in doing ϕ we could be simply assuming that there is a good reason in A's context that could help to sustain certain motivational shift (even if A's motives cannot be connected, by now, with the desirable features exemplified by doing ϕ). Thus, even we accepted that evaluations of rationality are somehow focused on A's attitudes, externalists would accommodate this general constraint by locating it into a wider framework composed by external reasons and by assuming that the function of our evaluations of rationality is primary formative in terms of sensibility or motivational habituation.

5. Assessing the *impasse* from a normative-based perspective

In the two previous chapters, I presented a general structure – or a general argumentative framework as I called it sometimes (chapter 5, section 1) - aimed to solve the *impasse* surrounding *Meta-ethics* (chapter 2, section 4.2). The structure tried to establish the psychological nature of our moral commitments by reflecting on the reasons-referring role of such commitments. Once the above structure was presented, I briefly summarized a set of argumentative alternatives to give content to the structure, and to secure a proper conclusion about the type of psychological state that is expressed by a moral opinion. Among the alternatives available to supplement or give precise content to the above argumentative framework, I stressed one in virtue of its wide appeal in recent meta-ethical debates. To many people, the argument above could be supplemented in the following way:

> (1) Necessary, to be under a moral obligation to ϕ is to have a good reason to ϕ
> [*Rationalism*]

> (2) Necessary, one/A has a good reason to ϕ only if one/A could be motivated to ϕ
> [*Reason internalism*]

(3) Necessary, if one/A is motivated to ϕ, then one/A must either desire to ϕ or desire to φ and believe that by φ-ing she will φ[*Humean theory of motivation*]

Therefore – by (1) to (3) – one A is under a moral obligation to ϕ only if either A desires to ϕ or A desires to φ and believe that by ϕ-ing he will φ.

When I presented this argument I did *not* argue for it at any stage. Instead, my main concern to put the argument on the foreground of the discussion was motivated by a meta-methodological interest. In essence, when I proposed it I had in mind a conditional sort of claim. I thought that if we were able to reconstruct the dispute around a certain premise of the argument (in my case premise 2) we might well conclude that no particular conclusion from this argument could be minimally warranted. As no agreement at the level of premise (2) can be sustained without avoiding disagreement, any appeal to use the argument above to overcome the impasse surrounding meta-ethics is misguided (Chapter 5, section 1 and 2).

Once this meta-methodological interest is noted, the last two chapters have been concerned with a single aim. I have tried to illustrate how the point assumed by (b) is essentially *disputed* or, to put in other words, how the alternatives that are available to fill the conceptual space suggested by the reasons-referring function of our moral opinions conflict among them, which cancels in the end any possible route to attain a straight and sound route to restart meta-ethics. To *illustrate* this negative point we need to show that, in fact, the assumption about the status of our reasons is essentially disputed. But how can we do such thing? How can we persuasively show that there is a deep disagreement about the second assumption - (b)?

To answer this question I started by delineating, in chapter 5, a maximally plausible or qualified version of reasons-internalism, attending further to two basic intuitions supporting this meta-normative thesis. After presenting such qualified version of reasons-internalism (W2') I tried to argue against it by taking two routes. On the one hand, I tried to show that the qualified version faces internal problems (section 2). On the other hand, I argued that, even if the qualified version of reasons-internalism were stable, we should reject the two basic concerns animating reasons-internalism (section 3). The combined effect of this wide-ranging rejection should impulse, or so I believe, a significant amount of scepticism on the prospects of reasons-internalism.

However, by rejecting the qualified version of reasons-internalism we do not have yet a sufficient basis to claim that the general argument is problematic, because one of the premises that it contains is underdescribed. We also need to eliminate a certain possibility. This possibility would say that even if reasons-internalism were problematic, we could motivate a case for externalism about reasons by assuming the negative image extracted from our very criticism of the internalist interpretation of the second assumption. But the thing is that if the prospects of reason-internalism are quite low, the prospects of reason-externalism as alternative meta-normative position are not much better. When we attend to the semantic and psychological proposals defended by reasons-externalism we might conclude that this meta-normative proposal is under several problems. I have suggested, for instance, that we do not even have a clear image of what is the content of our normative thoughts along an external line of interpretation. Also, I have noted that externalism cannot offer (as reason-internalism did at this point) a unified explanation of the motivational import of our normative assertions. At the end, I have defended that externalism is also problematic to conceptualize the status of our reasons. If what I have said above is plausible, then we should not assume that there is a knockdown argument to restart *Meta-ethics* that departs from the general structure going from (a) to (c).

Chapter 7

Restarting *Meta-ethics* from *Meta-rationality*

1. Introduction

2. Restarting *Meta-ethics* by embracing *Meta-rationality*

2.1 A general framework for *Meta-rationality*

2.2 Giving content to *Meta-rationality*

3. A first argument from *Meta-rationality:* Smith's argument

3.1 Unfolding Smith's argument

3.2 Evaluating Smith's argument

4. A second failed argument from *Meta-rationality*: Kalderon's argument

4.1 Unfolding *Kalderon's intransigence argument*

4.2 Assessing *Kalderon's intransigence argument*

5. Conclusions

Chapter 7

So far, I have argued for three basic claims:

(1) The *object* of *Meta-ethics* should be understood along with a set of platitudes that shape our moral stance. To the extent that these platitudes characterize our moral perspective, *Meta-ethics* should always take them into account to offer a coherent, systematic, and consistent second-order reconstruction of our moral experience. In offering such reconstruction *Meta-ethics* should be guided by an ideal of *accommodation* (chapter 1).

(2) Endorsing *accommodation* as a methodological ideal is not going to be free of consequences. Among these we could anticipate, by considering the evolution of this methodological ideal, a situation where two meta-ethical theories exemplify a similar degree of accommodation with respect to an identical set of platitudes about morality. Following Michael Smith, I have referred to this scenario as a methodological *impasse* (chapter 2).

(3) In assessing the *impasse* and possible responses to it, I have implicitly assumed that there are four methodological options available:

(i)   We might deny, if not the conceptual possibility of the *impasse*, at least its effective reality.

(ii)   We might simply accept the reality of the *impasse*, and embrace some sort of methodological scepticism about the overall prospects of meta-ethics.

(iii)   We could minimize the strength of the *impasse* by simply assuming that one theory will be, in fact, a more plausible methodological choice *all things considered -* because, let us say, it would avoid some undesirable practical consequences.

(iv)   We might well accept the reality of the *impasse* without being committed to any variety of scepticism on the prospects of meta-ethics.

I have endorsed (iv) as the best methodological option to confront the *impasse*. After that, I have proceeded to evaluate the relative strength of two common strategies to solve the *impasse* by appealing to two central contexts where the concept of *reason* is commonly used. First (chapter 3), I presented a classical argumentative schema that starts by assuming a set of motivational assumptions about our moral judgments to endorse a positive conclusion about the nature of the psychological state that express when we accept a moral opinion. After that

(chapter 4), I focused on some criticisms to that argumentative schema, trying to show how a basic assumption accepted by such arguments (*motivational internalism*) is not well secured, and does not license a knockdown conclusion about the psychological nature of our moral judgments. I have worked by persuasion at this stage: even if we have no knockdown criticism against such arguments, the simple fact that we have not reached any level of agreement about the interpretation of the modal premise about motivation makes the use any motivational-based argument to support a meta-ethical conclusion about the nature of our psychological states very dubious.

After that negative stage, I proposed (chapter 5 and 6) that the same negative strategy could be useful to reject the argumentative structure that highlights the psychological nature of our moral opinions by stressing some features associated with our normative sense of *reason*. I argued that as far as a certain *assumption* about the metaphysical status of our normative reasons cannot be the focus of an unqualified agreement, *any* particular argument containing a version of such type of statement to support a meta-ethical conclusion might not be a direct route to highlight the psychological nature underlying our moral commitments. In essence, by assuming the soundness of this type of arguments, we are simply relocating the nature of the impasse, passing it from meta-ethics to theory of reasons and rationality. Or so I argued until here.

The two next chapters will introduce another type of conceptual strategy to minimize the *impasse* surrounding contemporary *Meta-ethics*. In essence, I will assume that, insofar as moral opinions can be intuitively understood as discursive devices connected with a variety of normative practical reasons, the fact that we are able to ask a set of second-order questions about our judgments of reasons or rationality (semantic, ontological, epistemological, and psychological) should be considered as a methodological short-cut to obtain central answers about the nature of our moral opinions. I will refer to this methodological intuition, which stresses the importance of second-order issues of rationality for *Meta-ethics,* as **M:**

> **M** If we want to resolve a set of *meta-ethical* questions (focused, for example, on the psychological nature of the state that we express when we accept a moral sentence), then we should be concerned with some second-order questions about our reasons and rationality

By appealing to **M** as a general methodological intuition, I will be implicitly committed with two general, albeit derivate, methodological constraints:

(a) I will try to depict a straightforward argument from *Meta-rationality*, i.e. a *particular* argument whose directness and intuitive character would offer by itself sufficient basis to neutralize the intuitive appeal of the image locating *Meta-ethics* at the juncture of a methodological impasse.

(b) Since the sort of particular argument required to give content to **M** will be necessary located into a wider theoretical context (one where, for example, this appeal to questions about normativity and rationality is justified in terms of the attractiveness of *moral rationalism* as philosophical thesis) I will /I will need to support the proposed reduction of *Meta-ethics* to *Meta-rationality* with a further argument supporting *moral rationalism*.

Keeping in mind **M** along with the two further constraints expressed by (a) and (b), I will focus on the plausibility of some arguments that have stressed the importance of *Meta-rationality* to offer some answers about the second-order nature of our moral opinions. Michael Smith and Mark Eli Kalderon have recently offered two direct attempts to answer questions about the psychological state expressed by a moral opinion by turning our attention toward issues of *Meta-rationality*. I will devote a significant part of this chapter to comment both arguments.

2. Restarting *Meta-ethics* by embracing *Meta-rationality*

I will not simply presuppose that *Meta-ethics* would be methodologically developed by working on some topics related with the status of our norms of reasons or rationality. In this sense, rather than merely assuming that **M** could be a feasible option by itself, I would wish to appeal to a certain view in support of a methodological shift to *Meta-rationality***.** To do this, I

will be basically concerned with an argument delineated by Michael Smith in support of **M** [Smith, M. 2005. p. 8-9 and 21-22].

2.1 A general argument for *Meta-rationality*

Michael Smith appeals to three related intuitions to unfold the sense in which *Meta-rationality* could help to restart *Meta-ethics*. First, Smith accepts a commonsensical point. He notes that besides of being merely guided by norms, we are able to reflect on the nature of these norms. We all share, or so Smith assumes, a general disposition to reflect on our own norm-guided practices and performances beyond the deliberative or first-personal importance with which certain norms shape our deliberation. By means of this sort of capacity we are able to ask some basic questions about our norms – questions about their ontological reference, their meaning, about the sort of psychological scaffolding required to make them operative, or the evidence to be adduced to support them, etc… – to achieve an additional, theoretically-based understanding of our normative-guided behaviour that would help to improve our normative-guided performances [Smith, M. 2005. p. 7-8. See also Railton, P.2006a and Pettit, P. 2007 for two recent discussions of the general ability]. Smith notes:

> "Viewed from this perspective, meta-ethics is just one area in which we can raise the host of interrelated semantic, metaphysical, psychological, and epistemological questions (…) So the question is why we concern ourselves with these questions in meta-ethics as opposed to the similar questions that arise in meta-*rationality*, or meta-*law*, or meta-*etiquette*" Smith, M. 2005. p. 7

But if we can reflect, from a second-order perspective, on a variety of norms guiding our behaviour why, Smith asks, *Meta-ethics* has been usually understood along with a claim of methodological priority with respect to other meta-level approaches that focus on the status of other normative systems? He offers the following answer:

"One natural answer is that ethics purport to be, in some yet to be specified sense, a more basic normative system than the others. For example, the norms in each of these other domains, or the most important of them, might be thought to reduce to, inter alia, moral norms. If this were so, then it would be clear why we should begin by answering meta-ethical questions rather than meta-level questions in the other domains, for the only way we could fully answer the other meta-level questions would in that case be by first answering those same questions in meta-ethics" Smith, M. 2005. p. 7-8

However, as soon as we support the methodological priority of *Meta-ethics* in respect to other meta-level approaches on the basic status of our moral or ethical norms, we find a problem lurking around. It is simply *not* true that moral norms are more basic than *all* possible types of requirements. Norms of rationality, for instance, are supposed to be more basic than moral norms. Thus, if the basic assumption supporting the methodological importance of *Meta-ethics* is grounded on the supposedly basic character of moral norms, and if norms of rationality are more basic than moral norms, there is an obvious consequence for the status of *Meta-ethics*. Smith writes, in line with it:

"(…) if this is a good reason for supposing that we should answer meta-ethical questions prior to answering questions in meta-law, then a similar argument shows that we should answer meta-level questions about rationality before we answer meta-ethical questions". Smith, M. 2005. p. 8

So, the second intuition that Smith proposes in support of **M** says that *Meta-rationality* is methodologically prior to *Meta-ethics* in virtue of the more basic character of norms of reasons and rationality**.**

After this, Smith implicitly endorses a *third type* of assumption. By means of this assumption, Smith is able to relate *Meta-ethics* and *Meta-rationality* not merely along a comparative dimension, but also along a constitutive one. Smith argues that we can reduce

Meta-ethics to *Meta-rationality* because moral norms conform a subset of the norms of rationality. As a matter of conceptual necessity, there is constitutive link between our moral requirements and the reasons that we have in a given context - or alternatively: between the actions that are morally required and those that would be rational. *Rationalism* is a plausible conceptual intuition about the way in which moral norms guide our behaviour [Smith, M. 1994. p. 62]. But if *rationalism* is plausible, a direct consequence about the status of *Meta-ethics* could be endorsed:

> "(…) then, in order to say whether in making moral judgments we express beliefs or desires we must first know whether in making judgments about what there is a reason to do or judgments about what is rational to do, we express beliefs or desires" Smith, M. 2005. p. 8

Taken together, these intuitions state directly the content of **M**. However, it yet remains obscure how M should be understood. To put it loosely: How might we proceed if we wanted to restart *Meta-ethics* by endorsing a second-order approach in line with *Meta-rationality*? To answer this question, Smith delineates the following argumentative structure, in which he could locate his own proposal:

(1) Let us suppose *that Meta-ethics* were dealing with a psychological-oriented question, one referred to the type of mental state being expressed by us when we endorse a moral opinion [Gibbard, A. 2003. p. 6. Kalderon, M. E. 2005. p. 5-8]

(2) Let us suppose that moral considerations do in fact represent genuine reasons for action, i.e. let us suppose that moral considerations have the status of considerations that count for or against the agent's acting on a certain way [Scanlon, T. 1998. p. 18. 2007. p. 6. Wallace, R. J. 2006b. p. 187]

(3) Let us suppose that *Meta-rationality* were dealing with a two-level task: (i) *Meta-rationality* would try to define what kinds of concepts are more basic in the realm of reason and rationality; (ii) *Meta-rationality* would ask, from a second-level standpoint, about the kind of mental state expressed by our judgments of reasons or rationality

Therefore, if (2) were a plausible ground to reduce *Meta-ethics* to *Meta-rationality*, then we could give an answer to (1) in the indirect way pointed by (3.ii)

Evidently, the above argument is not free of problems. I will accept that (1) is a reasonable way to state the focus of a great deal of contemporary theorizing about the nature of morality. In a sense, there is a sort of agreement around the possibility to formulate meta-ethical questions not as directly referring to the nature of moral properties but rather as trying to reach to a sound image about the nature of our moral thought. In asking the latter question, a clear response about the type of psychological state that plays a dominant functional role in our expressions of moral commitments would be essential to respond to classical questions about the status of morality – questions about its ontological import, the sort of evidence required to justify our moral claims, or the meaning of our moral statements [Gibbard, A. 2003. p. 7-9. Sayre-McCord, G. 2006. p. 57-60. Joyce, R. 2006. p. 51-52. Ridge, M. 2007. p. 52 for three recent statements stressing a psychological-oriented view of *Meta-ethics*]

But even if we could give sense to (1), the main problem with the methodological program depicted by the previous argument would be to make sense of premise (2) and (3). In a certain way, the strongest assumption supporting the conditional contained in the conclusion is surely expressed by premise (2), the one claiming for *rationalism*. The whole argument aimed to reduce *Meta-ethics* to *Meta-rationality* does not work if *rationalism* is false. Evidently, whether *rationalism* is true or not is a large issue and I do not want to settle it down here [see Wallace, R. J. 2006b]. However, I would like to say something more precise about the sense of *rationalism* I am referring to all along this discussion. Additionally, I would like to defer on two arguments in support of *rationalism* to precise what I am assuming here by using that label.

About the first task, the obvious thing to note is that *rationalism* is an old philosophical term. Its philosophical pedigree can be found in some classical disputes around the nature of the psychological abilities involved in our moral judgments. In this well-known sense, *rationalism* stands for a simple claim: moral truths can be apprehended by the exercise of our *rational* capacities, i.e. by the exercise of the capacities involved in the recognition of truth and falsity. As we all know, Hume rejected this precise sense of *rationalism*. In Hume's famous dictum, "reason alone can never be a motive to any action of the will" [Hume, T. 1978. 413]. In traditional sense then, *rationalism* is basically a claim focused on the status of

those capacities that allow us to act in accordance with certain duties or obligations. These capacities, Hume argued, cannot be identical to our rational capacities, i.e. they cannot be equated with a certain type of *beliefs* [See Watson, G. 2004/1973. at page 16-17. Bricke, J. 1996. p. 14].

But this is not the sense of *rationalism* I have in mind here. In my sense, *rationalism* is mainly a thesis about the type of normative import underlying any system of requirements. Focusing by now on the moral case (see Chapter 8. Section 1 and 2), *rationalism* accepts that

[Moral Rationalism] If A morally ought to ϕ then A has a reason for ϕ-ing[99]

If this meta-normative principle is true, anyone who is saying about herself something like 'I ought to ϕ' is licensing our saying of her that 'She has a reason to ϕ'. Equally, if I say 'you ought to ϕ' and you ask me why you ought to ϕ, I will give you a bad answer if I simply say (increasing my tone) that you must ϕ. Surely, a better answer in this context would provide you with a reason to ϕ, i.e. with a consideration that would justify your ϕ-ing.

An important caveat about the type of the considerations offered as reasons is in place here (See Chapter 1, section 3.2). In my favoured sense, when I facilitate you a reason to justify your ϕ-ing, I am neither merely saying that you have a *moral* reason to ϕ (I am not assuming that the very fact that ϕ-ing is morally right is giving you a reason to ϕ) nor presupposing that your reason to ϕ is grounded on the possibility of punishment for not ϕ-ing. Rather, what I am suggesting is that when you ought to ϕ there be a set of considerations that would justify your ϕ-ing. These considerations cannot be ignored when you decide what to do in this precise context. And, crucially, they are not constitutively dependent of the existence of an institutional context of sanctions and benefits. These considerations (the value we exemplify at the moment we are responding to them) are conforming the normative pressure to which we normally appeal to when we offer a justification of your ϕ-ing in terms of reasons [Scanlon, T. 1998. p. 19. 95-97. Joyce, R. 2002. p. 42-44. Wallace, R. J. forthcoming][100]

---

[99] Wallace describes *rationalism* in ethical theory as the view holding that "moral considerations are reasons for actions". Wallace, R. J. 2006. p. 2. This is Gibbard: "On a narrow conception of morality, in contrast, moral considerations are just some of the considerations that bear on what it makes sense to do". Gibbard, A. 1990. p. 41. And, finally, Scanlon's sense: "If it would be wrong for a person to do X in certain circumstances, then he or she has strong (normally conclusive) reasons not to do so". Scanlon, T. 2007a. p. 6

[100] Dancy writes: "(…) the point that verdictive judgments do not contribute to the situations on which they pass judgment is only an application of the more general truth that thin moral concepts cannot be used to add to the

Once this point is more or less clear (see Chapter 8, section 1 and 2 for an additional treatment), we can focus on the second task I announced before. At this stage I will defer on Smith himself (see Chapter 3, section 3). If Smith is right, it would be possible to defend that *rationalism* is a direct consequence of the most reasonable way to state the *practicality* of our moral judgments by assuming the following argument [Smith, M. 1994. p. 80 and 2004/1997. p. 284-287]:

(i) (***Practicality's claim***) As a matter of *conceptual* necessity, if A believes that he is morally required to ϕ then A will be motivated to ϕ-ing, *if A is rational.*

(ii) (***Rational expectative***) As a matter of conceptual necessity, a rational agent will be motivated to ϕ-ing if he believes that he is rationally required to ϕ.

Therefore, if A believes that he is morally required to ϕ then – by (i) and (ii) – A believes that he is rationally required to ϕ

Therefore, to be morally required is to be required by reasons[101]

---

store of reasons. That an action is good, or right, is no reason to do it. It is the features that make the action good or right that are the reasons for doing it, and to say that it is good or right is merely to express a judgment about the way in which other considerations go to determine how we should act". Dancy, J. 2004. p. 16. The same point can be found in Raz, J. 1978. p. 12. Cuneo, T. 2007. p. 65 and Scanlon, T. 2007b. p. 6

[101] There is an alternative formulation of this argument offered by David McNaughton and Piers Rawling (McNaughton, D. Rawling, P. 2004. p. 114-115):

(1) Rational agents will do what they judge they are morally required to do
(2) Rational agents will judge truly
     Thus
(3) Rational agents will do what they are morally required to do
     But
(4) Rational agents will do what there is (most) reason to do
     Hence
     (R) There are reasons to do as morality requires

Therefore, as far as it seems that we could appeal to some kind of movement to support the second premise contained in the general argument for *Meta-rationality*, the main problem for Smith consists of giving content to the very stance of *Meta-rationality*. How could we precisely state a sense for *Meta-rationality* beyond the quite empty appeal for a second-level approach concerned with some fundamental issues about reasons and rationality?

2.2 Giving content to *Meta-rationality*

In order to fully understand how *Meta-ethics* could be reduced to *Meta-rationality* we should define more precisely what *Meta-rationality, understood as a second-order stance, could be*. We might be able to point, so to speak, to the sort of problems that should be the focus of *Meta-ethics* once it were reduced to a second-level inquiry about the nature of our reasons and rationality. I will suggest here that there are at least three candidates to give content to *Meta-rationality*. Let me locate them into Smith's general argument in support of *Meta-rationality*:

(1')   Let us suppose that *Meta-ethics* were dealing with a basic sort of question, referred to the type of mental state expressed by us when we endorse a moral opinion

(2')   Let us suppose that moral considerations do in fact represent genuine reasons for action, i.e. having the status of considerations that count for or against the agent's acting on a certain way

(3')   Let us suppose that *Meta-rationality* were dealing with a two-level task: Let us suppose that *(i) Meta-rationality* were trying to define what kinds of concepts are more basic in the realm of reason and rationality and that *(ii) Meta-rationality* were asking, from a second-level standpoint, either about **(a)** the type of psychological state expressed by us in deciding or planning what to do, about **(b)** the type of psychological state being expressed by us in the cases in which we ascribe a reason to a system or, finally, **(c)** about the psychological processes underlying situations where A and B disagreed about the appraisal, in terms of reasons, of a certain outcome or action.

Therefore, if (2) were a plausible ground to reduce *Meta-ethics* to *Meta-rationality*, then we could give an answer to (1) by favouring either *(ii.a), (ii.b),* or *(ii.c)*

Once Smith's argument is unfolded along these lines it would seem as if (once secured the rationalistic intuition and accepted that meta-ethical questions are questions about the type of psychological state expressed by a moral opinion) we could have three possible routes to pursue meta-ethical questions. Let me present them in turn.

a. *Meta-rationality* could study the psychological processes and capacities required for non-moral decision and deliberation

A possible way to fill the methodological space left open by Smith when he argued for *Meta-rationality* would be to simply assume that, accordingly to *(ii.a)*, in embracing *Meta-rationality* as a methodological program we should be concerned with the role played by certain normative concepts in basic decisional contexts, i.e. contexts where we are deciding, deliberating or planning about *what to do*. Although several philosophers are currently endorsing a quite similar methodological approach to meta-ethics – one concerned with the second-order nature of our decisions and deliberative processes - the most clear and sustained instance of this methodological strategy is due to Allan Gibbard [See Hussain, N. Shah, N. 2006. Enoch, D. 2007. Bratman, M. 2007. Introduction, especially on page 11, for some approaches stressing the deliberative standpoint in getting meta-ethical conclusions].

Gibbard introduces his meta-ethical view by accepting three basic points. First, he claims that we should shift the focus of attention in current *Meta-ethics* from ontological or semantic questions - i.e. those concerned with the metaphysical or the semantic nature associated with our basic normative and moral terms - to a more restricted interest in characterizing the identity of the *psychological* state being expressed by a normative or an ethical statement [Gibbard, A. 2003. p. 6 and 183].

Second, he further endorses an assumption about the proper scope linked to this psychologically oriented question. In essence, Gibbard insists on the importance of focusing on *simple normative terms* in order to answer the question about the type of psychological states being expressed by us in contexts that involve more complex normative concepts [Gibbard, A. 2003. p. 7].

Finally, Gibbard argues for a shift on the focus of *Meta-ethics*. He defends that, among the basic normative uses, we should start by asking for the sort of state being expressed by us in those cases where the basic normative terms are located in *deliberative contexts*, for example,

when planning or deciding what to do[102] [Pettit, P. 2006. p. 1083. Bratman, M. 1987]. Gibbard denies, in essence, that the practical decisions that appeal to the basic normative concept *the thing to do* (ones located into larger patterns of plans and beliefs of the agent) can be understood as expressing directly a belief by themselves (either on a natural or a non-natural sort of content). However, as he points at several places, the concept expressed by a predicate as *the thing to do* is "remarkably fact-like in its behaviour" [Gibbard, A. 2003. p. 5. See also page 4].

For this reason, Gibbard assumes throughout his book that the concept expressed by those situations where we use the predicate *the thing to do* works as an *expressive* device, i.e. as a device aimed *not* to *represent* a realm of facts (whether they are understood as natural, non-natural, psychological, non-psychological, etc…) but rather to *express* or to *avow* a certain state of mind, a certain intention or, as in Holmes's case, a certain decision. In Gibbard's sense, to conclude that ϕ-ing is *the thing to do* is simply to conclude what to do [Gibbard, A. 2003. pp. IX-X, 8, 13 and 77. Pettit, P. 2006. p. 1091[103]]

Thus in Gibbard's account it would be possible to interpret *(ii.a)* as assuming a quite specific conditional claim. It could be understood as defending that if you were trying to do *Meta-ethics* by previously illuminating some (second-order) questions about reasons and rationality, then a feasible option would be to ask directly about the type of *normative psychology* that we require to think and plan *what to do*. Once we reach a response to this question (and by no means it will be an easy task to obtain an uncontroversial answer to the sort of psychological framework required by our planning nature [Pettit, P. 2006. p. 1095 and 1099-1103]) we would assume that the same type of normative psychology presupposed by us to explain how these basic decisions or intentions can be sustained (an expressivistic psychology, if Gibbard is right) could help to understand the way we deploy various normative concepts in other, more complex, contexts of deliberation.

Evidently, Gibbard's methodology is extremely ambitious. In an obvious sense, we could accept the methodological principle he endorses - the one stressing the importance of simple decisions to offer a basis to construe an extended meta-ethical treatment of normative terms -

---

[102] Note that hereafter *the thing to do* will be understood as an application, after a deliberative process in the agent has reached its end, of the concept *is the thing to do*

[103] Pettit writes: "Gibbard's view, then, is that although thinkers-planners can reason, availing themselves of the concepts of what is the thing to do, or what is okay to do, they need not instantiate any states other than decisions or plans that might have been there in the absence of such concepts; they need not instantiate any new representations by virtue of exercising the concepts" Pettit, P. 2006. p. 1094

without necessary committing ourselves with any meta-ethical view extracted from simple instances of decisions. In a certain sense, even if a methodological program along the lines stressed by Gibbard were persuasive, more need to be said to secure the transition from basic cases of decision to complex normative scenarios As Bratman notes:

> "We build step-wise from the simpler to the more complex (…) we can try to see to what extent models of strong forms of agency can be constructed in a way that is neutral with respect to debates in meta-ethics between cognitivists and expressivists views; though (…) we need to remain alive the possibility that such neutrality will not be available at the end of the day" Bratman, M. 2007. p. 11 [104].

Leaving these worries aside, the thing to keep in mind by now is that Gibbard's account (or something along these broad conceptual lines) could be used in order to give sense to the idea that *Meta-rationality* can secure *Meta-ethics*, as *(ii.a)* affirms.

b. *Meta-rationality* could study those situations where we normally ascribe normative predicates to systems

Besides of the intimate link that connects the use of normative concepts with the expression of a decision, normative concepts are usually found in non-deliberative contexts, far from our immediate first-personal perspective. I have now in mind, for example, those contexts in which we evaluate others' behaviour in terms of rationality or reasonableness without deciding ourselves what would be the most reasonable or rational thing to do. That we usually assume such instance is a platitudinous fact. But if so, then surely any favoured second-order account of the status of our moral concepts should pay attention to this fact.

---

[104] Clearly, the two steps could be independent, not implying, of course, that we are always in position to appeal to a definite psychological framework to sustain a meta-ethical account of deliberation. David Enoch writes, for example: "Thus, in deliberating you commit yourself to there being (normative) reasons relevant to your deliberation. Now, this sense of commitment need not entail an explicit belief that there are such reasons (…)" Enoch, D. 2007. p. 38. As I pointed before, Parfit (a robust meta-normative realist himself with whom Enoch agrees at several places of his paper) directly assumes, in turn, that normative *beliefs* are about "irreducible normative truths" See Parfit, D. 2006. p. 339 and 342-345

When we pay attention to the *attributive* sense attached to these concepts, a central fact emerges: the rules determining the use of our concepts of *reason* and *rationality* are constrained by descriptive properties of the system to which we apply these terms. Any second-order account must assume that when an agent ascribes a normative predicate to an intentional system, he is usually resting on certain descriptive properties of the system [Foot, P. 2001. Chapters 1-2 and Thompson, J. J. 2007].

Thus, if a non-cognitive psychology was needed to explain some assumed platitudes about the role of our basic normative concepts when they are embedded in a first-personal context of deliberation, surely we also need to introduce a cognitive psychology to explain how we are able to ascribe normative predicates to agents and intentional systems in a socially structured context, i.e. in situations of common interaction. I will say more about this option when I present Smith's argument. By now let me introduce the last natural option to give content to (ii) in the argument above.

c. *Meta-rationality* could study instances of polarized normative disagreement

In those circumstances where A is asserting that *p* while B is asserting that not-*p* we all assume that A and B *disagree* and that their disagreement is about *p*, i.e. about whether *p* is the case, whether *p* is true or, equivalently, whether *p* can be asserted with guarantees in the context in which A and B are located [Stevenson, C. L. 1944].

To resolve such cases of disagreement two basic strategies are usually appealed to. Either we suggest that both, A and B, should attempt a further assessment of the evidence available in order to decide whether *p* can be asserted [Ayer, A. J. 2001/1936. Boyd, R. 1988] or, alternatively, we check, by means of questions or tests, A and B's capacities of deliberation, trying to eliminate the explanatory hypothesis suggesting that the disagreement between A and B could be explained by some kind of incoherence, epistemic shortcoming, or irrationality affecting to A or B [Brink, D. O. 1989. Chapter 7]. Unfortunately, it is quite common that even when we confront A and B to these further tests we find that they remain disagreeing about whether, in a same evidential context, *p* can be asserted. These are cases of deep or *polarized disagreement*.

Cases of *polarization* are common in many domains. For instance, you could say that a system S is behaving in a rational way while I claim that S is behaving in a severe irrational

way – even when I am being responsive to the same amount of evidence than you – when the evidence is not limited to external features but also to S's desires, beliefs, dispositions and so on. Leaving aside the easiness by which we can find normative disagreements mirroring factual disagreements, what would these situations imply in terms of the current debate? What could we learn about *Meta-ethics* by focusing on these cases of polarization or disagreement about reasons?

Mark Kalderon has recently proposed an obvious response to this question [Kalderon, M. E. 2005][105]. Let us take, Kalderon argues, some cases of *polarized disagreement*, and circumscribe them to cases of polarized epistemic disagreement, i.e. cases where *p* refers to a factual proposition and A and B disagree about whether *p* is true even when A and B are exposed to the same evidence and both are equally reasonable. Then consider the different processes and reactions that A and B will show when confronted with the fact that they are disagreeing about *p* even when (i) *they acknowledge a similar amount of evidence* relevant to whether *p* can be asserted and even when (ii) *they assume that both were equally coherent and rational*. Now, let us suppose that A and B were exemplifying a certain process R. Let us suppose, for instance, that in cases of epistemic disagreement along with (i) and (ii), A and B are normally disposed to subtract some probability to their relevant attitudes when they are aware that they are disagreeing about *p*. In this case, we should assume that R is a typical reaction to cases of polarized epistemic disagreement, i.e. cases whether two agents are trying to reach to a verdict about whether a certain belief could sustain or justify a certain assertion.

Now, let us suppose that we find a bit later to A and B disagree about some normative issue *q* (legal, moral, etc). In this case, A and B not only disagree about *q* but also about what is the relevant evidence fixing whether *q* is the case (because of the fact that both (i) and (ii) are true of them). They are disagreeing, so to speak, about the reasons that could support a normative assertion. Now, let us suppose that neither A nor B are disposed to discount a certain level of credence to their beliefs (*q* and *not-q*) as R suggested in the case of epistemic disagreement. Once this point is reached the question is obvious: Could we say that A and B are believing in a certain content (that *q* or that not-*q*) if they are *not* exemplifying the same processes (R) exemplified in cases of epistemic disagreement? In what sense do they believe a moral or legal assertion if they are not disposed to behave as if they believed that a certain moral or

---

[105] I was firstly acquainted with Kalderon's proposal by attending a talk delivered by Jussi Suikannen at the Department of Philosophy of Reading University in April of 2006

legal propositional content is the case? Evidently, some people assume that, as far as A and B are not revising their evidence in line with (R), they are not disagreeing about an epistemic issue, i.e. they are not believing, in any sense, that *q* or that *not-q*.

I will present later this general strategy in a more detailed way. The important thing by now, however, is to grasp the potential importance that our understanding of those cases of genuine or polarized epistemic disagreement could have to reach central conclusions about the nature of our normative disagreements on practical issues. In sum, by comparing the processes underlying cases where we are weighing reasons (epistemic versus practical) we could obtain an independent perspective on the sort of state expressed in a situation of normative disagreement and, by extension, in any moral dispute. It would be by doing meta-rationality that we could obtain meta-ethical insights.

These are, in sum, three available routes to ground an approach to *Meta-ethics* based on second-level issues of rationality. Evidently, there is a sort of continuity between these possible options for *Meta-rationality* and those contexts where we normally use our basic normative concepts (reasons, rationality, reasoning). The *guiding* nature or the decision-focused character of our normative concepts (ii.a), the *attributive* function usually associated with our normative concepts when they are embedded in social networks (ii.b) and, finally, the way in which normative concepts are shaped by cases of *disagreement* (ii.c) offer, from my point of view, three possible ways to ground *Meta-rationality.*

In a certain sense, if we needed arguments able to provide support for a certain second-order thesis about the nature of our moral judgments – once assumed that *Meta-ethics* is facing a certain methodological *impasse-* we should look for direct or simple arguments coming from *Meta-rationality*. In my view, as far as properties as *directness* or *simplicity* are not equally shared by the three previous accounts, we should favour the approaches to *Meta-rationality* with a direct and simple argumentative line. Thus, I am going to argue that (ii.b) and (ii.c) should be preferred over (ii.a) in virtue of the more neutral character and the less number of assumptions that they endorse compared to (ii.a).

3. A first argument from *Meta-rationality*: Smith's argument

In this section I want to focus on an argument departing from the perspective I have been referring here as *Meta-rationality,* partially sketched by Michael Smith. His argument reaches substantive conclusion about the type of mental state expressed by our moral commitments. By reflecting on our attributive practices – by means of which we ascribe a particular normative predicate to a system (i.e. *rational*) - we should be able to infer the psychological nature of the abilities involved in these situations. [Smith. M. 2005. p. 23-25. Pettit, P. 2006].

There are three things to keep in mind about Smith's argument. F*irst,* this argument works by assuming that *moral rationalism* is a well-grounded thesis. Once this point is accepted, Smith assumes that by reflecting on the second-order nature of the cases in which we apply our basic normative terms we can obtain important information regarding the type of psychological state expressed by our moral opinions [Smith, M. 2005. p. 8-9].

The *second* thing to note is related with the *force* of Smith's argument. What I am going to present here is a version of what Smith himself has referred to as *a flat-footed attempt* to argue for cognitivism. Evidently, if the argument is introduced by Smith in such vein, we should be cautious about the status of the argument - at least if we want to derive a sound conclusion from it. Probably, the expression pointed above works as a sort of conversational proviso, which tries to make explicit that the argument is not endorsed, as it stands, by Smith himself and that in consequence it is expressing a certain general stance toward our moral opinions shared by cognitivists in general but not by Smith's own position [Smith, M. 2005. p. 22].

Finally, the *third* point to note is somewhat related with what I have just claimed. If Smith's argument can be taken, in principle, as a *flat-footed attempt* to ground *Meta-ethics* on *Meta-rationality*, its force *as an argument* will depend on the lack of any powerful criticism on it. In essence, even when the force of Smith's argument can be questioned in advance, as far as the natural alternatives to the *flat-footed argument* are not very persuasive, Smith's argument will be more credible than we had imagined previously.

3.1. Unfolding Smith's argument

To formulate the argument, Smith assumes that we have reached to a basic normative concept by means of a sound analytic route. He notes that *a psychology that meets all the requirements and ideals of reasons and rationality* is the most basic normative concept. Smith notes that by means of such concept we could define other normative concepts such as *ought* and *good*. Let us refer to this definitionally basic concept as B [Smith, M. 2005. p. 21].

Smith notes that we could isolate, by means of a process of analysis of certain actual and possible cases where B is normally applied, a set of features or platitudes compounding something close to a *network analysis* of B, a *mature folk-theory* about B, or a *reference-fixing description* of B [Smith, M. 1994. p. 29]. Smith notes that we could isolate certain core features in any possible functional network around B to fix its content. Let us assume, for example, that this subset of features – extracted from several possible situations where B is consistently applied - refers to the *coherence*, *unity*, and possibly the *informedness* of a psychological set. Let us refer to these features as the subset [C, U, I].

Finally, Smith claims that we should be able to grasp consistently a set of non-normative features associated with those situations where we successfully apply B to a system. We usually grasp these features [C, U, I] in contexts where we have a set of descriptive features.

Once these preliminary points are accepted[106], Smith's argument for cognitivism departing from the level of *Meta-rationality* goes as follow [Smith, M. 2005. p. 22]:

---

[106] Smith will assume a version of *moral functionalism* to identify the precise descriptive properties that moral terms stand for. *Moral functionalism* can be identified by means of a three-stage schema. To identify the properties moral terms stand for:

- We first need to establish a set of *a priori* truths about our moral terms, a set of commonplaces or platitudes such that "anyone who knows how to use the term *fair* is in position to see that they hold" [Jackson, F. Pettit, P. 1995. p. 193]. These platitudes are formulated by using evaluative and non-evaluative predicates [Jackson, F. Pettit, P. 1995. p. 192].

- Once these truths are obtained we should get, at a *second stage*, a general framework to establish the *meaning* and the conditions under which any given moral *property* is instantiated. So, if we considered the case of *rightness*, moral functionalists assume that the *meaning* of *rightness* would be equivalent to the functional role occupied by the term *right* in the network of platitudes conformed by our moral terms [See Jackson, F. Pettit, P. 1995. p. 195. Jackson, F. 1998. p. 131]. Accordingly, the property that *rightness* stands for is equivalent to the functional property associated with the rightness-role when we consider the network of platitudes or commonplaces that encapsulate the meaning of *right* [Jackson, F. Pettit, P. 1995. p. 194].

- Finally, we should assume that, as far as *supervenience* is a sound conceptual claim, the functional roles by means of which a given property is identified must be paired off with a

(1)   B is a basic normative concept.

(2)   *A priori* we can establish that B = [C, U, I].

(3)   To argue for whatever ideal instantiation of the set [C, U, I] is to argue for the *a posteriori* significance of some non-normative features over others.

(4)   *A priori*, any psychological system instantiating the non-normative features on which the set [C, U, I] supervenes will be one on which we can apply B (by 2 and 3).

(5)   When we predicate a property based on certain non-descriptive features about an object, we are expressing a minimal belief: *the belief that a certain object is instantiating the non-normative features.*

Therefore, by (4) and (5), when we make a basic judgment of reason or rationality over a system we are expressing a belief: *the belief that a certain system is instantiating the non-normative features on which the set* [C, U, I] *supervenes.*

(conjunctive) descriptive property. The property that *rightness* stands for will be equivalent to the (conjunctive) descriptive property on which the rightness-role is supervening [Jackson, F. Pettit, P. 1995. p. 194. Shafer-Landau, R. 2004. p. 70. Dancy, J. 2006. p. 123-126].

Any functional account of the meaning and reference of evaluative predicates is subject to two well-known problems. The first one is *metaphysical*. In an intuitive sense, when we assume that *rightness* is the role-filling property occupying the rightness-role in given functional network, we are subject to a well-known ambiguity about the metaphysical import of the property identified by means of the functional analysis. In essence, by assuming that the property *rightness* stands for the descriptive property on which the rightness-role supervenes, we could imply either that *rightness* is the very *ground-level property* on which the rightness-role supervenes or, alternatively, that rightness is the *second-order property* of having the descriptive property (whatever it is) that occupies the rightness role. See Jackson, F. Pettit, P. 1995. p. 197. Scanlon, T. 2007a. p. 7. Pettit, p. 2007. p. 230-231.

Additionally, a related *epistemological* concern should be faced once we endorse functionalism in any domain. In essence, while to believe that an action is right could be taken as equivalent to believing that [a given action would have the property of having the property that occupies the rightness role], other people would assume that to believe that a given action is right should be equivalent to believing that [a certain action would have the descriptive, role-filling disjunctive property on which rightness is supervening]. Which option, among the alternatives just noted, would more adequate to grasp the metaphysical status and the epistemic nature of our moral or evaluative beliefs is an open question.

Evidently, the argument appeals to the functional account previously sketched to explain how a certain variety of meta-ethical cognitivism would be able to pair off evaluative and descriptive properties in a precise way. But there is an important difference: if in the previous context functionalism was used to support a certain claim about the nature of moral properties (that moral properties are functional properties), in the present context functionalism is used by Smith to get a substantive and direct conclusion about the type of state expressed by a moral opinion. In terms of this alternative appeal to functionalism, once we have accepted that functionalism is expressing a very plausible view about the way we fix the meaning of any concept (normative concepts being included here), the fact that only by appealing to meta-ethical cognitivism we can make sense of the way a functional schema evolves in simple cases of normative attributions should be understood, or so Smith argues, as a powerful point in support of a cognitive meta-ethical framework. In sum, although functionalism is a common ground in arguments *assuming* meta-ethical cognitivism and arguments that argue *for* meta-ethical cognitivism, and although precisely for this reason some people could claim that an argument as the one above presupposes a cognitivists stance, the real fact is that in both types of contexts the functionalist premise is being used as a neutral tool to specify a plausible way to determine the meaning or content associated with certain basic concepts. But leaving aside this point, how plausible is Smith's *argument*?

3.2. Evaluating Smith's argument

Smith notes that there are two routes to attack his argument. While the first critical route will be concerned with the tenability of premise (4), the second one will attack premise (5). Let me start by Smith's case against premise (4). This premise assumes, remember, that we could apply, as a matter of *conceptual* or *a priori* necessity, a B-predicate to a certain agent by merely considering how a set of commonsensical features, ones that could fix the reference of B, are exemplified by certain non-normative or descriptive features – by certain psychological features. However, once we are aware of the fact that there is a deep disagreement about whether a given set of non-normative or psychological features support, in fact, the application of a B-predicate to a certain system, the force and directness of premise (54?) is put under fire. But how can this be possible?

The truth is that these cases are not only possible but also quite common. Smith offers some examples coming from theory of decision, and cases of disagreement can be found even at more basic levels. For example, consider how some people defend that the conditions stated by an instrumental requirement of rationality could fix by themselves a necessary and sufficient criterion to use the term *rational* while some other people assume that an agent violating a certain *categorical* requirement of instrumental rationality could be described as rational [Smith, M. 2005. p. 22 and 24 for the conception of imperative offered here and Kelly, T. 2006]. Smith notes that to dismiss this disagreement we should either to accept that those discussing this issues "aren't really as smart as they seem to be" [Smith, M. 2005. p. 23] or that the argument that would bring the agreement among those disagreeing about the criteria of applicability of a certain normative term is to come. As neither of the possibilities is very plausible (because people working on this issues are very smart and because we have been exposed to too many arguments), Smith concludes that what premise (4) states cannot be true, i.e. that to move from a set of descriptive features to the application of a B-predicate is *not* an a priori secured step. The natural outcome will be that:

> "(…) the only credible conclusion for us to draw is that the move from an specific claim about the way our psychology is structured to the conclusion that psychology is one that meets all rational requirements and ideals of reasons is not a priori, but is rather a matter of decision. We make the specific claim and then, if we have the relevant kind of desire that people have psychologies that are so structures, we express our desire in the form of a judgment to the effect that psychologies of that kind meet all rational requirements and ideals of reasons" Smith, M. 2005. p. 23

But Smith does not accept this conclusion, offering, in turn, a sort of *reductio*. He claims, in essence, that if the above argument were accepted, then we should conclude that every domain where we find a situation of disagreement along the lines mentioned above (one where smart people have been working for a long time to a sound argument) would fall under the same sort of non-cognitivist description. Evidently, as far as it would be implausible to

suppose that the second-order interpretation of several discursive domains actually involved in disagreement would assume non-cognitivism, Smith concludes that the argument cannot be sound, at least as it stands.

Smith will offer, nevertheless, a second route to reject the argument going from (1) to (6) by focusing on premise (5). At this stage the criticism will be quite familiar. In terms of it, as far as any judgment saying that a certain psychology is falling under the scope of a B-predicate can be associated, as a matter of conceptual necessity, with a certain motivational response from those who endorse the judgment, a judgment containing a B-predicate cannot consist simply in a belief – a belief of the type expressed by (5) - because beliefs, under a well-known story, are motivationally inert [Smith, M. 2005. p. 23-24]

Smith rejects the objection by presenting, again, a *reductio* working on those cases where we propose that a certain *desire* should be in place to secure the motivational link between a given application of a B-predicate, a certain action-type ϕ, and a certain motive to ϕ. Let us suppose, Smith adduces, that we would need a desire (*a desire to have a psychology that meets all requirements of reasons and rationality*) to be motivated to ϕ when we intend φ and believe that if φ then ϕ. If non-cognitivists were right, the desire just mentioned would suffice to motivate us to ϕ in presence of the previously noted psychological states. However, by assuming such *desire* it is far from clear that we would be able to directly comply with the previous instrumental requirement:

> "The person we have imagined must therefore not only desire to have a psychology that meets requirements and ideals of reasons, but he must also believe that, since he desires a certain end and believes that that a particular means is a means to that end, so having a desire for that means would give him a psychology that meets requirements and ideals of reasons, and then…and then what? The possession of this particular desire and means-end belief is not enough to guarantee that the person we are imaging desires the means to this particular end either (…) how is the putting together of this particular desire and belief accomplished? If we suppose that the person

must have a further desire, then we are off on an infinite
regress" Smith, M. 2005. p. 24

Smith will conclude that as far as an infinite regress is lurking once a certain desire is introduced to explain cases of normative motivation (*a desire to have a psychology that meets all requirements and ideals of reasons*) such desire could not be able support, by itself, the motivational link in a reasonable way, invalidating, in sum, the non-cognitive criticisms on Smith's argument coming from the motivational realm [see also Railton, P. 2004a and 2004/1997].

Once Smith reaches this point, he must either reject the motivational intuition on which the objection is working (*motivational internalism*) or, alternatively, he must offer an alternative explanation for the internal link between our normative judgments and our motives. Evidently, anyone familiar with Smith's meta-ethical views could imagine which option he proposes at this precise stage. In a nutshell, he argues that we could explain the intimate link established between our normative judgments (the ones which ascribe a basic normative predicate like B to a certain action) and our motives without appealing to the direct role played neither by a desire nor by a belief. We could explain the motivational connection established between our normative assertions and our motivating reasons by simply assuming a sort of *disposition to coherence*. Such capacity, although located beyond any particular psychological state such as belief or desire, would be able to secure a reliable pairing between our evaluative beliefs and our motives. Thus, neither beliefs nor desires – although, again, beliefs enable in Smith's conception the work of the capacities just noted – are directly relevant to explain the motivational reliability of our normative claims [Smith, M. 1994. p. 177. 2004/1997. p. 267-269.2004/2002. p. 351 and 2004b].

Thus, as far as Smith's favoured solution to the motivational problem posited by non-cognitivism could be considered, in principle, consistent with a cognitive-based approach to the content of our normative beliefs, the non-cognitivist attack presented above will not harm the argument going from (1) to (6). The consequence is that, as Smith announced, as far as we do not have a knockdown criticism against *the flat-footed argument* we should accept it.

4. A second argument for *Meta-rationality*: Kalderon's *intransigence argument*

The current section will further explore another argumentative possibility associated with the methodological path facilitated by *Meta-rationality*. I will be concerned here with an additional proposal that delivers a straight response to the question about the psychological nature of our moral opinions by asking a second-order question about our judgments of reasons and rationality. In essence, while we were concerned before with the second-order nature of those cases where we ascribe a basic normative predicate (*rational*) to a system, in this section I will focus on certain second-order features associated with those situations where two agents disagree about the reasons supporting a certain claim. I will assume, following Smith's general intuition connecting meta-ethics and meta-rationality, that if we were able to obtain a conclusion about the type of thought being expressed in these basic normative contexts then (in virtue of the conceptual link facilitated by *rationalism*) we would be able to get a positive basis to answer some basic questions about the second-order nature of our moral opinions.   To explore this alternative route, I will present an argument recently developed by Mark Kalderon:

(1) In contexts of *polarized cognitive disagreement* - i.e. contexts where we disagree with *epistemic peers* about whether to *accept* a factual or theoretical sentence on behalf of others - we are under a *lax obligation* to inquire further into the grounds supporting our acceptance [Kalderon, M. E. 2005. p. 19 and 25][107].

(2) Moral acceptance is a variety of acceptance on behalf of others [Kalderon, M. E. 2005. p. 26-27].

(3) If *moral acceptance* were a variety of *cognitive* acceptance, we should expect that in those contexts of *polarized moral disagreement* an agent would be under a lax obligation to inquire further into the grounds supporting his acceptance of a moral sentence. [Kalderon, M. E. 2005. p. 26-27 and 34]

---

[107] See Kelly, T. 2006 for the use of the expression "epistemic peer" in the context of an informed discussion on epistemic disagreement.

(4) In those contexts of a *polarized moral disagreement* an agent is *not* falling under a lax obligation to inquire further into the grounds for his moral acceptance [Kalderon, M. E. 2005. p. 35]

Therefore – by (2) to (4) - we should conclude that moral acceptance is *not* an instance of cognitive or factual acceptance [Kalderon, M. E. 2005. p. 37].

Evidently, in order to fully grasp the force of this argument we should further highlight several aspects of it:

- First, it should be made explicit the general sense we are giving to *acceptance* and the way in which *acceptance* on behalf of others in contexts of polarized factual or cognitive disagreement is associated with a duty of a certain type (premise 1)

- As far as the very possibility of comparison between the duties involved in a moral and a factual examples of disagreement is intimately connected with the existence of a structural similarity among domains in terms of the type of *acceptance* involved, we should further justify why *moral acceptance* is a variety of *acceptance on behalf of others* (premise 2).

- It should be further explained why, while cases of *polarized factual disagreement* exemplify a certain lax epistemic duty, this very lax obligation is not operative in those cases of *polarized moral disagreement* (premises 4).

- Finally, insofar as the centrality of premise (3) is accepted, it should be required to look into the conditional claim expressed by this premise even if the previous points were secured. So, a last step required to clarify the argument should move around the grounds supporting premise (3).

Once these points are noted, let me begin to pay close attention to Kalderon's argument.

4.1 Unfolding Kalderon's *intransigence argument*

(a) Acceptance, factual acceptance on behalf of others, and polarized disagreement (premise 1)

First, let me say something about what *acceptance* is. In general, in recent philosophical discussions *acceptance* has been understood as a particular type of doxastic stance having a separate proposition or a set of propositions (a theory) as focus. So, in the same sense in which we can occupy a belief-stance, a hypothetical-stance, a pretence-stance, or even a fancy-stance toward a certain propositional content, we could *accept* that a certain thing is the case [Gibbard, A. 1990. Chapter 4. Bratman, M. 1999/1992 Velleman, J. D. 2000/1992]

One of the most important questions that *acceptance* exemplifies once it is introduced along a line of commonality with other doxastic attitudes is, evidently, the sense in which *acceptance* would be comparable or reducible to other doxastic attitudes or, to a set of doxastic attitudes. Clearly, among the plurality of questions found at this level, one of the most pervasive concerns is whether *acceptance* and *belief* are comparable in epistemic terms or whether, in a related sense, *acceptance* can be reduced to a certain type of *belief*. This precise topic (whether *acceptance* and *belief* are comparable epistemic states and whether they can be mutually reduced) has been so pervasive in recent literature that, leaving aside a precise answer about the type of relationship established between them, we have ended by getting a much better understanding of the nature of *belief* and *acceptance,* and how the latter is connected to certain psychological and practical dynamics of our agency [See Bratman, M. 1999/1992 for this pragmatic understanding of *acceptance*].

Among the many features appealed to when *acceptance* and *belief* are compared, three basic points are customary noted. It has been defended that while belief is an involuntary stance, not subject to direct control by the agent, a state of acceptance could be, to some extent, *voluntary* attained by the agent. Further, it has been defended that while genuine belief is immune to any instrumental use (because, for example, we cannot believe as a means for some end that we are pursuing – at least in full awareness), we can easily find instances of acceptance whose rationale or justification is purely *instrumental*. Finally, it has been assumed that while the criteria determining whether we can occupy a belief-stance should be

considered as context-independent, the criteria determining whether we can accept a given sentence are essentially *context-dependent* and quite sensitive to the nature of the domain in which we are accepting a sentence [See Railton, P. 2006a for these features].

Although these central points (*voluntary control, instrumentality,* and *context-dependence*) would surely constitute a central part of any understanding of *acceptance*, in presenting his argument Kalderon stresses a different set of features about *acceptance*, which relate to the fact that some instances of acceptance are sometimes located into a *public* context of inquiry. Before presenting these public features of acceptance, let me note some minimal points about how Kalderon understands *acceptance* itself.

In Kalderon's sense, there are two essential points to secure our claim that A is accepting a certain sentence S about a certain domain of discourse M (biology, history, morality, physics, etc…). The *first* feature assumes that, in accepting S, A "no longer takes himself to have reasons to investigate further, to continue to inquire whether or not to accept the sentence" [Kalderon, M. E. 2005. p. 6]. As result of this basic platitude, the *second* feature assume that A will be disposed to rely on S in further theoretical and practical reasoning and, importantly, that A's relying on S could be justified in virtue of certain norms. In Kalderon's sense, when a sentence S is accepted by A in any given domain of inquiry two types of norms are relevant in explaining A's state of acceptance of S:

 (A) *Internal norms of acceptance* that specify what kind of evidence or which would justify the acceptance of a sentence belonging to M. *Perceptual* support, *truth*, explanatory *coherence*, or *fidelity* to certain sacred book could be considered, in Kalderon's sense, as internal norms of acceptance working on factual, scientific, and religious domains [Kalderon, M. E. 2005. p. 14-15]. According to Kalderon, internal norms have a set of secondary features worth mentioning:

- At the general level, these norms govern *static patterns* of attitudes by means of principles that require, in a deontic style, a certain attitude in virtue of a certain other patterns of attitudes [Kalderon, M. E. 2005. p. 16].

- As consequence of the peremptory status associated with internal norms of acceptance, a failure to obey such norms would constitute a sufficient basis to secure an accusation of *irrationality* [Kalderon, M. E. 2005. p. 18]

- Although internal norms give voice to a strict requirement over an agent's possible attitudes, they are susceptible to be overridden by the pressure of other type of norms, so that internal norms of acceptance at t_1 cannot determine by themselves the attitudes an agent will show at t_2 [Kalderon, M. E. 2005. p. 6, 16, 12-13 and 18].

- Importantly, as far as acceptance can be exemplified *in a public domain of inquiry,* internal norms should fix public-dependent norms of acceptance. The reasons or the evidence required to accept a given sentence in a public domain of inquiry would be dependent upon the access of others to such reasons or evidence. In essence, the reasons we adduce in accepting a sentence in a public context should be reasons *on behalf of others*[108].

(B) *External norms of acceptance* that constitutively regulate any type of inquiry External norms are usually formulated by means of a set of epistemic ends that regulate the activity of inquiry (the type of epistemic goals Kalderon cites include the avoidance of inconsistency, the lack of incoherence or contradiction*,* and the re-evaluation of our evidence in certain situations, among many others). *External norms* are characterized by sharing a set of additional features:

- First, the epistemic ends that give substance to those *external norms* express *lax requirements* (requirements that the agent has certain discretion to ignore) about how the particular obligations derived from them should be usually weighed against other requirements that constrain our attitudes and our behaviour in some particular contexts (practical, prudential contexts).

---

[108] Kalderon writes: "Whereas acceptance for oneself is both the object and the ground for individual inquiry, acceptance on behalf of others is the object and ground of public inquiry: if an competent speaker accepts S on behalf of others, he takes himself to have sufficient reasons to end public inquiry about S" Kalderon, M. E. 2005. p. 23. He writes later: "In uttering 'Abortion is permissible' (…) he accepts 'Abortion is permissible' on behalf of others for he takes to himself to have access to a grounding reason that is a reason to accept that sentence not only for himself, but for everyone else as well" Kalderon, M. E. 2005. p. 35

- Additionally, external norms range, in Kalderon's sense, over a *temporally* extended pattern of attitudes and they are say that, if at t_1 you are in P, at t_2 you ought $_{lax}$ to be in Q.

- Finally, as a direct consequence of the contextual discretion associated with these lax requirements, we can ignore them in some cases without incurring in any strict charge of irrationality[109].

The importance of this distinction goes as follows. Kalderon refers to *external norms* to explain how we should proceed, in epistemic terms, if we were located in a context of *polarized factual disagreement.* In Kalderon's sense, in a *polarized disagreement* about a factual matter, two agents - A and B - disagree about whether to accept a given sentence S (i.e. 'that the Spanish government should have kept its troops until the end of the hostilities') only if the following conditions are met:

(a) A and B share the same amount of relevant evidence about S

(b) A and B are both rational and equally coherent in accepting the evidence with independence of whether they accept that *S* or that not-*S*

Kalderon assumes that when A and B are disagreeing about *S* under the above description, they are not only disagreeing about whether *S* is the case (about whether *S* can be accepted) but rather about what are the relevant normative principles that determine the acceptance of *S* [110]. Thus, in some cases of disagreement between two agents, once we have fixed certain evidential parameters shared by these agents and once we have assured a certain parity among the agent's capacities of reasoning and rationality, the only plausible way to explain their

---

[109] Kalderon writes: "Suppose that Edgar accepts certain sentences that are indirectly inconsistent (…). Suppose that Edgar comes to recognize that the sentences he accepts are indirectly inconsistent. Perhaps Bernice has explicitly given him the argument leading to direct inconsistency. (…) While Edgar may be obligated as an inquirer to adopt the end of resolving the inconsistency, it would not be irrational for Edgar to persist in his acceptance, saying to Bernice: 'That's really interesting, I'll have to think about that later, but right now I have to pick the kids" Kalderon, M. E. 2005. p. 17.

[110] Where a normative principle in the factual domain is understood in terms of the reasons that are relevant to accept a given sentence.

disagreement should appeal to the fact that they may be working with different normative principles, which stress different features as normatively significant in a shared realm of evidence.

However, a caveat is in place here. By merely accepting that (a) and (b) can be useful to describe certain cases where A and B disagree, we cannot infer that the disagreement can only be explained by appealing to different normative frameworks for A and B. It would be possible, or so I believe, that A and B were previously committed with non-identical sets of beliefs or - if you are happy speaking in terms of degrees of belief - that their set of beliefs were subject to different prior probabilities. If this were soothe case, the fact that A and B disagree about S could not be *only* explained by appealing to different normative principles shared by A and B. Rather, that they disagree about a certain issue in a context where (a) and (b) are true of A and B could be explained by the fact that their backgrounds beliefs are not identical neither in substantive nor in probabilistic terms [Harman, G. 1986. Chapter 3. Christiansen, D. 2005 Chapter 1]

This point should motivate a modification in the formula above. By taking into account the phenomenon just noted, we should defend that A and B are involved in a polarized disagreement if and only if:

(a*)   A and B share the same amount of relevant evidence about S.

(b*)   A and B are both rational and equally coherent in accepting the evidence with independence of whether they accept that *S* or that not-*S*.

(c)*   A and B share an identical set of backgrounds beliefs (or an equal distribution of prior probabilities).

Let us suppose that these sorts of situations were plausible or, at least, imaginable. How should we react to them? If we considered this situation from a *detached* standpoint, couched in terms of the permissibility associated with a certain instance of acceptance, the two logical responses available when each agent is faced with this sort of disagreement (persistence and revision) might be equally permissible once certain contingent features associated with the agents' situation are assumed. For example, it could be permissible for A to persist on his acceptance of S if this persistence were related with a general policy of A favouring epistemic

conservatism. Equally, it could be permissible for B to revise his non-acceptance of S if a debunking explanation were offered to him in order to explain why he disagrees with A about S[111]. In these situations, a *claim of equal permissibility* would be operative, which would allow A to persist in the acceptance of S or to revise the acceptance of S if certain conditions were given [Kalderon, M. E. 2005. p. 14-15]

An interesting feature emerges, however, when we consider the cases of disagreement from a *non-detached* standpoint, i.e. from the perspective of each agent involved in a case of polarized factual disagreement. When we consider these cases of disagreement from an *internal* perspective, an additional feature should be added. Now, A and B disagree about S only if:

(a)   A and B share the same amount of relevant evidence about S.

(b)   A and B are both rational and equally coherent in accepting the evidence to accept that *S* and that not-*S,* respectively.

(c)   A knows that (a) and (b) and B knows that (a) and (b).

Once the situation is viewed from this personal perspective, Kalderon assumes that both agents (A and B) are under a *lax obligation* to inquire further into the grounds of their respective acceptance. In terms of this lax obligation, once it is assumed that A and B are concerned with the truth of S in a context along the lines sketched by (a)-(c), they are required to re-examine their evidence for accepting or rejecting S. Inasmuch as Kalderon understands

---

[111] I will understand *permissibility* here as referring to the normative space left by the direct application of basic or primary internal norms in a given domain. Permissibility is also ruled by norms, and in fact Kalderon sometimes is close to suggest that these are a subset of the internal norms referred above. In Kalderon's sense, these norms enumerate a large set of domain-dependent conditions that secure the permissibility of a certain attitude leaving aside the validity of previous norms of acceptance. Once the combination of sentences required by an internal norm is in order, claims of permissibility point toward certain possible alternatives. Kalderon offers the following examples:

> *(Doxastic Conservatism)* If A has fulfilled the internal norms in M and A believes S and A has no positive reason to reject S, then it would be permissible for A to retain S

> *(Perfect Symmetry)* If A has fulfilled the internal norms in M and A believes S, and A and B are occupying a symmetric stance with respect to the relevant evidence for S, and A knows this fact, then it would be permissible for A to revise S

See Kalderon, M. E. 2005. p. 14-15 and 22

this duty as an instance of an *external norm* of acceptance, the set of conditions I noted when I introduced external norms could be applied to these situations [Kalderon, M. E. 2005. p. 17-19 and 21]. For example, although this *lax requirement* expresses an inescapable epistemic obligation (in the sense that any instance of public inquiry couched in terms of truth will fall under the scope of this obligation) it does not support a *deontic obligation* on us. Kalderon assumes that other norms extracted from other activities in which we are involved, could impose opposed constraints, which would preclude the fulfilment of these external norms [Kalderon, M. E. 2005. p. 18].

Besides, although if we accepted the point just noted we should assume that A and B are *not* being irrational if they are *not* disposed to re-examine their evidence in these cases of disagreement, Kalderon defends that A and B are lacking some kind of *epistemic merit* because they persist in their particular commitment without being disposed to revise their grounds either to assert S or to reject S. Thus, although a *lax obligation* favours a certain attitude as a consequence of the external norms regulating our inquiry, such obligation does not secure a charge of irrationality.

But the important question is, evidently, why are A and B under this type of obligation? Why are they required to revise their evidence in a context of disagreement along (a)-(c)? In Kalderon's sense, the fundamental ground supporting this *lax obligation* to reconsider is connected with the *public* character of the acceptance that A and B exemplify. In essence, Kalderon assumes that, insofar as the sort of reasons that A and B appeal to in a public context of acceptance are *reasons on behalf of others*, a disposition to reconsider will show the best available response to the type of reasons A and B face in a public context of acceptance - ones that should/could be accepted by themselves and by others [Kalderon, M. E. 2005. p. 15, 19, 22-23 and 25]

Having said that, in a case of *polarized factual disagreement*, A could be under a *lax requirement* to revise his prior acceptance of S if he is under (a), (b), and (c) and he exemplifies an instance of acceptance on behalf of others. This is what premise (1) affirms in a quite convoluted way [. Kalderon, M. E. 2005. p. 18-19].

(b) Moral acceptance is acceptance on behalf of others (premises 2)

In order to support premise (2) Kalderon should prove that the norms governing moral acceptance are a subset of those norms governing the cases of acceptance on behalf of others. He argues for this claim assuming a certain view about the authority of morality. [Kalderon, M. E. 2005. p. 27]

Let us take any moral sentence ('abortion is wrong') as it is usually uttered in a context of moral discussion. To connect moral acceptance and acceptance on behalf of others, Kalderon starts by assuming that any instance of acceptance moves around a basic normative concept, the concept of reason. By assuming the concept of reason as given, the moral authority can be highlighted in an easier way. Any instance of moral acceptance conveys, or so Kalderon assumes, a reason to act in a certain way supported by some reasons, which are tracked by certain normative principles. Let us denominate the first role that reasons are playing in moral acceptance 'the convey-role', and refer to the second type of role played by reasons in contexts of moral acceptance as 'the grounding role'. Kalderon notes that when reasons play 'the convey role' in any given instance of moral acceptance, the authority of morality is understood along two well-known phenomena:

a. ***Precedence***: Moral sentences convey reasons that either override or cancel any conflicting non-moral reasons available in the circumstances to determine what to do [Kalderon, M. E. 2005. p. 28]

b. ***Noncontingency***: Moral reasons conveyed by moral sentences are not contingent upon the speaker's acceptance [Kalderon, M. E. 2005. p. 29]

On the contrary, when reasons play 'the grounding role', the authority of morality is related to a very important dimension:

c. ***Well-groundedness:*** the reasons grounding the utterance of moral sentences are applicable not only to the speaker at the moment of the utterance but to everyone else as well [Kalderon, M. E. 2005. p. 31]

. Besides, by exemplify a grounding role in this precise sense, moral reasons are able to explain a basic conversational demand linked to the utterances of the moral sentences of a competent speaker:

> d. **Demand**: In uttering a moral sentence, a competent speaker demands that his audience accepts the uttered moral sentence [Kalderon, M. E. 2005. 32]

Once it is accepted that these features conform (along with the conversational demand) the core of our notion of moral authority, Kalderon assumes that they can be only explained if moral acceptance were an instance of acceptance on behalf of others [Kalderon, M. E. 2005. p. 33]

(c) The conditional claim (premise 3)

As I have said at the beginning of this section, Kalderon defends that a possible route to answer some meta-ethical questions about the type of psychological state expressed by us when we accept a moral opinion, should focus on how a certain *regulative disposition* (D) is exemplified in some contexts of factual or cognitive acceptance, in which we disagree with other agent about whether a given factual content can be asserted. Once the regulative disposition is identified in the context of *factual* acceptance (a lax obligation to reconsider), Kalderon assumes that we could arrive to a certain meta-ethical conclusion about the psychological nature of our *moral* acceptance by asking whether D is taking place/present in those cases in which we disagreeing with other agent about whether a moral commitment can be asserted. In Kalderon's sense, if D were present, we could defend a commonality between the type of state we express in cases of moral acceptance and the state we express in cases of factual acceptance. On the contrary, the lack of D in cases of moral disagreement would suggest that *moral* acceptance is not based on the same type of mental state we express in cases of *factual* acceptance. I will refer to this general intuition as *the conditional claim*:

> *Conditional claim*: If moral acceptance expressed a cognitive process, then, in a context of moral disagreement about reasons a person should be under a lax obligation to inquire further into the grounds of his acceptance [Kalderon, M. E. 2005. p. 38]

Evidently, once the conditional claim is assumed we need to know if we are required, in cases of moral disagreement, to further inquire on the grounds of our acceptance. Premise (4) offers precisely an answer to this question.

(d) Responding to (polarized) moral disagreement (premise 4)

In cases of polarized *moral* disagreement framed in terms of reasons, Kalderon claims that we are *not* under a lax obligation to re-examine our grounds for moral acceptance, i.e. we are not required, in the lax sense, to inquire further into our reasons as it happens, on the contrary, in cases of polarized factual disagreement [Kalderon, M. E. 2005. p. 35]. Kalderon's argument at this stage is conceptual in spirit. In Kalderon's sense, that we are *not* under a lax obligation to revise the grounds of our moral opinions in contexts of polarized moral disagreements can be established in virtue of how certain reactions (our lack of embarrassment at the personal level if we do not reconsider our moral commitments and the lack of criticisms over our behaviour from a third-personal stance even if we were persisting in our moral opinions) implicitly assume the non-applicability of a norm that would requires the revision of the grounds of our moral commitments.. In sum, premise 4 offers a positive basis to highlight the psychological nature of our moral commitments through the application of an instance of the conditional test depicted above. As far as we are not under a lax obligation to reconsider in cases of moral disagreement, we are not expressing a cognitive state in such contexts. Or so the conditional claim implies.

4.2 Assessing Kalderon's *intransigence argument*

Once every premise of the argument has been highlighted, how plausible is it? And further: How good is it as an instance of an approach to meta-ethics based on meta-rationality? Let me remind you again the overall structure of Kalderon's argument:

(1) In a context where we are polarized about whether to accept a *factual* sentence *on behalf of others* we are under a *lax obligation* to inquire further into the grounds that support our acceptance.

(2) *Moral acceptance* is a variety of acceptance on behalf of others

(3) If *moral acceptance* were a variety of *factual* or *cognitive* acceptance, we should expect that in the contexts of *polarized moral disagreement* an agent would be under a *lax obligation* to inquire further into the grounds that support his acceptance of a moral sentence.

(4) If we attend to our conceptual practices, is not the case that in the contexts of a *polarized moral disagreement* an agent is under a lax obligation to inquire further into the grounds for his moral acceptance

Therefore – by (1) to (4) - we should conclude that moral acceptance is *not* an instance of cognitive or factual acceptance.

The basic problem faced by Kalderon's argument is at the level of premise (1)[112]. When it is conjoined with premise (2), both imply premise (3) – what I took to be the crux of Kalderon's argument. For this reason, in order to resist the conditional claim expressed by premise (3) we should reject either premise (1) or premise (2). As I am happy to accept premise (2) – because of the fact that I am sympathetic with the general account of moral authority that premise (2) implies - premise (1) will be my target in what follows. Although we could question this premise from three possible routes, I will only use one critical standpoint. Nevertheless, let me suggest how the other routes should proceed in order to resist the conclusion posited by (5)

First, it could be simply argued that what premise (1) assumes cannot be assessed neither as true nor as false. If this were so, premise (1) should not be considered as a good candidate to support any argument. Remember that premise (1) is claiming that, once located in certain ideal conditions along the lines from (a*) to (b*) – an agent who accepts a sentence on behalf of others exemplifies a certain disposition to reconsider that is expressing by itself the

---

[112] Kalderon suggests that the overall argument could be resisted at premise (3), by endorsing a standard version of moral relativism. In terms of such meta-ethical theory, the truth-conditions of any instance of moral acceptance are relative to a certain moral framework. So, when A disagrees with B in a context of polarization, in uttering 'x is wrong' A is expressing a belief whose truth-conditions are relative to A's moral framework while that B (in uttering 'x is right') is expressing a belief whose truth conditions are relative to B's moral framework. If it were so, the fact that neither A nor B are motivated to revise their beliefs (even when they are able to acknowledge their symmetry in terms of evidence and rationality) should be explained by a basic relativistic assumption that they share (both assume that their moral perspectives are different). If relativisms were true, A and B could be expressing a belief even when they did not revise the evidence supporting their beliefs when faced with an agent endorsing an opposed moral opinions. And this possibility is precisely the negation of premise (3).

existence of a duty to reconsider operating in such context. The problem I am stressing here is that, as far as the conditions of (1) are extremely idealized, we cannot have any intuitive clue neither of the type of dispositions of an agent in such context nor, consequently, of the nature of our reactive attitudes toward these possible reactions. As far as Kalderon assumes that it is by means of the dispositions and expectations we have toward other's behaviour that we are able to grasp a duty in a certain context, a context of extreme idealization does not offer a very good basis to identify such duty [see Kalderon, M. E. 2005. p. 12 for a comment on the role of idealization].

However, even if we accepted the feasibility of the idealization proposed by premise (1), a problem would remain: How could we decide whether an actual instance of disagreement is exemplifying a valid instance of the idealized type depicted by premise (1)? After all, it would not be very clear how a duty inferred from a *maximally idealized* context of disagreement (one in the line depicted by premise (1)) should be used to identify the psychological state expressed in a *non-maximally idealized* context of disagreement (one in the line depicted by premise (3)) by asking if the same type disposition found in an *maximally idealized* context of disagreement (a disposition to reconsider the grounds of our acceptance) could be found in a *non-maximally idealized* context of disagreement. If the contexts depicted by premise (1) and premise (3) cannot be compared, by hypothesis, in terms of the degree of idealization that they support, then a quite reasonable stance toward Kalderon's argument would assume that even if we found a substantive difference in terms of the dispositions exemplified in contexts of type (1) with respect to the dispositions exemplified in contexts of type (3), both types of dispositions could be pointing, in the end, to the very same attitude. *As far as we do not precisely know how the difference in idealization could affect to our intuitions* about the type of state being expressed in each context we cannot infer any strong conclusion by comparing these two contexts: It would be possible, *as far as we know*, that when we judged the behaviour of an agent in an *idealized* context of factual disagreement, we tended to ascribe him a disposition to reconsider while that, when this judgment takes place in *a non-idealized context* of moral disagreement we tended to ascribe the agent a disposition to reconsider, even when the agent could be expressing a belief in both contexts. But if this point is plausible, Kalderon's argument is not conclusive at all. And we could have a second route to reject it.

But as I pointed above, I am not going to argue along neither of these lines here. On the contrary, I believe that a *third* type of criticism on premise (1) is more plausible.

Let me remind you, first, the norms that are involved in accepting a certain sentence S. In considering *acceptance*, Kalderon defends that some norms - *internal norms* – are able to determine, given a particular discursive domain D, which evidence should be adduced in order to fully accept a sentence S belonging to D[113]. Importantly, Kalderon assumes that the internal norms of acceptance associated with any given D could be classified either along a substantive criterion (relative to the subject matter of the domain itself: biology, history, medicine, astrology, theology, etc…) or along a less substantive one (stressing the degree of *publicity* associated with the acceptance of sentences in a given domain) [Kalderon, M. E. 2005. p. 14-15].

On the contrary, other norms governing acceptance - *external norms* - are domain-independent or, so to speak, domain-transcendent. *External norms* regulate, in Kalderon's sense, "the ends involved in *inquiry*. They represent the requirements on the ends to be adopted in changing one's epistemic state over time" [Kalderon, M. E. 2005. p. 16]. If I am understanding correctly Kalderon's distinction here, the norms that regulate externally our inquiry should be understood as referring to general epistemic or quasi-epistemic desiderata being applied to any instance of acceptance. In this sense, 'avoiding inconsistency', 'getting instrumental coherence', 'attaining a temporally consistent pattern of preferences' and, importantly, 'revising our evidence in cases of disagreement about a public issue' are all external norms in Kalderon's sense. All of them are involved in any instance of acceptance, leaving aside the content of the acceptance itself.

Kalderon defends that we have certain discretion to ignore external norms, which is absent for internal norms. This discretion should be explained, in part, because the external norms regulating our inquiry could conflict with other practical or prudential norms. In this sense, by not complying with an external norm that conflicts with other norms, we are not directly subject to a charge of irrationality - i.e. we are not irrational by not avoiding incoherence on a given case because by doing so we could crash? with some instrumental means to a certain practical end of us [Kalderon, M. E. 2005. p. 17]

Once I have noted the basic difference between *internal* and *external* norms, we should agree on the fact that premise (1) in Kalderon's argument establishes, *first*, a certain *internal*

---

[113] In an important sense, besides fixing the grounds required to positively accept S, internal norms should also be able to determine certain criteria to fix whether persistence or revision of our acceptance of S should be permitted given certain factors. As noted, as far as internal norms require a certain attitude from an agent, his failure to comply with such norms could support a charge of irrationality on him.

condition to accept a given sentence when we are in a public domain, i.e. when we are accepting a sentence on behalf of others:

> (i) A is accepting S on behalf of B – where S is a factual statement – only if A has positive reasons to accept S[114] and such reasons cannot be undermined by reasons accepted by B [Kalderon, M. E. 2005. p. 24 and 26]

Besides, in terms of an *external norm*, premise (1) establishes a concrete advice by saying that:

> *(j)* If an *undermining* relation were established between A's and B's reasons in respect to S's acceptance by A, then the best response for A to accept S on behalf of B would be to inquiry further on the grounds of S's acceptance to discover, if he can, reasons to accept S that B could not reasonably reject [Kalderon, M. E. 2005. p. 25]

The first assumption gives voice to a *substantive* thesis about the character of our evidence (our reasons) in contexts of public inquiry. The second assumption expresses, in turn, a normative *advice* (a *lax obligation*, an *external* norm) focused on any instance of public acceptance on behalf of others in contexts of disagreement [Kalderon, M. E. 2005. p. 24-25]

Evidently, (j) is a direct consequence of (i). This is so because once the *internal* norms governing the acceptance of S in a public context are defined in terms of the absence of any undermining factor, a direct consequence in those cases where A and B disagree about the reasons supporting the public acceptance of S will be that either A or B will have a duty to revise the available evidence (to neutralize any possible undermining factor.

And here comes my point. Although I defend a close connection between (i) and (j), I am opposed to the particular interpretation of acceptance favoured by (i) and, in consequence, to the particular advice offered by (j). In fact, although I can accept that certain public features of acceptance usually constrain instances of first-personal acceptance, these public-dependent features do not determine, by themselves, a divergent epistemic criterion to justify accepting a sentence. But (i) argues precisely for this point; it assumes that in contexts of public inquiry we are justified to accept a sentence only if a certain relational and public-based criterion (absence of any undermining reason) is the case. Thus, I will assume hereafter that the

---

[114] See Kalderon, M. E. p. 25 for this clause

specific features associated with contexts of public acceptance do not require any appeal to the absence of undermining factors to accept a sentence. I will refer to this alternative approach as *the non-divergence approach*.

> **The non-divergence approach**: The justificatory role of the evidence required for a person to accept a sentence in a public context is *not* conditionally dependent upon the absence of any undermining evidence coming from the evidential perspective of other agent in the same context of public acceptance.

The non-divergence approach offers an alternative interpretation to the kind of evidence internally required to accept a sentence in a public domain, i.e. an alternate reading of (i). In addition, *the non-divergence approach* accepts a tale about the duties and obligations of an agent located in a polarized disagreement. In terms of this negative stance, (j) should be rejected by favouring the following claim instead:

> **Non-divergence duty**: An agent in a context of public inquiry surrounded by disagreement is not required neither to look for a conclusive reason supporting his acceptance nor, alternatively, to explain the undermining evidence offered by other agent in case that he disagrees with him, in a polarized sense, about a factual sentence.

In sum, by endorsing *the non-divergence approach* we are rejecting both, a particular account of the kind of evidence required to accept a sentence in a public context - (i) - and, further, a particular reading of the sort of normative advice that could be applied in cases of polarized disagreement - (j). Once these points are assumed, in what follows I will be mainly concerned with the viability of the advice expressed by (j) in light of my favoured account. I will do so because if the advice expressed by (j) were ill motivated then (as far as (j) is giving voice to the commendatory side of premise (1)) we would not be under any lax obligation to reconsider the grounds of our acceptance in cases of idealized factual disagreement, as (1) is claiming. And ff premise (1) were false, premise (3) could not be true as it stands either, as the conditional claim expressed by premise (3) is entailed by both premise (1) and (2). Thus if premise (3) were not grounded at all - because of the falsity of premise (1) - then premise (3) could not be used to test the type of dispositions we exemplify in cases of moral

disagreement. And Kalderon's argument would be severely impaired. To contextualize these abstracts claims, let me refer to an example used by Kalderon himself:

> "Suppose that Bernice asks Edgar where the UCL Philosophy Department is and he says that is at 19 Gordon Square. Suppose, however, that Bernice has seen a flyer announcing that the Philosophy Department has moved from that address but she cannot now remember the 'new' address. Bernice would then accept a reason that undermine Edgar's acceptance of that address. Thus, Edgar would not be justified in accepting the address on behalf of others, because others, who are otherwise rational and reasonable, informed, and interested in inquiring about the address, accept undermining reasons and so reasonably reject that address (…) While Edgar would not be justified in accepting the address on behalf of others, he would be justified in accepting that address for himself" Kalderon, M. E. 2005. p. 24

Let me assume that Kalderon's example works as it stands. It should be assumed that both, Bernice and Edgar, are equally reasonable and rational and are informed and interested on the address of the Philosophy Department. Surely, the fact that Edgar has been living in London for a long time would facilitate a certain degree of deference by Bernice. But the thing is that, even once we have acknowledged this feature, Edgar and Bernice are sharing, or so Kalderon argues, an equal amount of evidence relevant to know where the Philosophy Department is and an equivalent degree of reasoning-ability and rationality to use this evidence to reach a conclusion about this particular issue. Once this context is established, Kalderon will claim that any instance of acceptance by Edgar and aimed to facilitate the relevant and public information (the address of the Department of Philosophy) should be qualified as such by the a further constraint: that Bernice is not accepting, at the same time, *undermining evidence*[115] against Edgar's acceptance of a given address. This is the point stated by (i).

---

[115] Understood as evidence that could invalidate the justificatory relation on which Edgar's utterance is based with respect to his own evidence

But Bernice saw a flyer announcing that the Philosophy Department has moved and this fact offers her a sufficient reason to undermine Edgar's acceptance, imposing a lax duty on him to revise his own evidence. So, if Edgar were interested in accepting a sentence on behalf of Bernice – if he were interested, for example, in agreeing with Bernice to meet at the Philosophy Department – Edgar should be under the obligation (once Bernice has endorsed undermining evidence against Edgar's acceptance) either to revise his evidence to explain away the undermining factor pointed by Bernice or, alternatively, to look for conclusive evidence in support of his own acceptance. And, again, this is so because Bernice would not consider the information encapsulated in Edgar's utterance as a sufficient basis to structure her practical deliberation (to coordinate with Edgar about where to meet, for example) unless she were lacking any reason to doubt the evidence supporting Edgar's utterance. This point, roughly, is stated by (j).

But the overall image is quite counterintuitive. Let me suppose that the previous case were susceptible to be analyzed in terms of *deference*. In an intuitive sense, as far as deference is conceptually connected with authority, the fact that Edgar should be able to weight a commonly available amount of evidence about a certain issue (the location of the Philosophy Department) in a more reliable way than Bernice could justify Bernice's deference on Edgar about such issue. So, in a certain way, we could accept that Bernice is deferring on Edgar because, although both share an identical body of evidence and a similar degree of rationality, Edgar is more familiar than Bernice with the context where the inquiry about S is embedded.

Once Kalderon's example *is constructed as a case of deference*, however, the role of a lax obligation to reconsider in cases of factual disagreement is less clear. And this is so because, as I am understanding *deference* here, a basic feature linked to such phenomenon is that the person on which someone has deferred about a certain subject matter is *not* obligated, even in a lax sense, to further inquire on the grounds of his acceptance unless a positive reason can be offered against his particular assessment.

Let me explain this point. I am claiming that if B defers on A about S, and B further points to A that a certain feature could undermine A's acceptance, a lack of disposition in A to revise the grounds of his own utterance (either by looking for conclusive evidence in support of his previous acceptance or by further counter-undermining B's evidence) could be perfectly in place. And it would be so because B deferring on A conceptually implies that A has authority to weight the evidence with some degree of independence with respect to B's access to the

available evidence. So, in a context where A and B disagree, if B is deferring on A, A is accepting S and B is supposedly presenting undermining evidence against A's acceptance of S, A should be disposed:

- To express his own evidence for S to B in a clearer way - if A could do such thing.

- To show B the reliability of his current evidence for S by explicitly referring to evidence shared by both being directly implied by A's acceptance of S.

But the important point about these alternatives is that, from the fact that the dispositions express intelligible responses to the peculiar deferential structure exemplified by Kalderon's cases, it should be established that a disposition to not reconsider could be an intelligible response in such cases (in which an agent defers on another agent who claims to have undermining evidence against the sentence accepted by the deferred agent). Rather, in contexts of disagreement that can be analyzed as cases of deference, any disposition among the ones pointed above could be a better response for A than the disposition to reconsider his own evidence that Kalderon favours when he assumes (j).

In sum, once accepted that premise (1) could be rejected through a more focused attack on (j) above, two consequences related to the status of Kalderon's argument can be assumed:

(a) Once we accept that (j) simply expresses a particular way to state the duty noted by premise (1), by rejecting (j) we also reject premise (1). So, even if the ideal situations exemplified by premise (1) were plausible (or conceivable, or imaginable) we could resist the whole argument by assuming that in certain contexts of factual disagreement analyzed in terms of *deference* others dispositions beyond reconsideration could be intelligible and, in the overall, advisable.

(b) Once premise (1) is put under fire, the conditional hypothesis expressed by premise (3) is not able guide in the identification of the nature of the psychological states expressed by us in cases of moral disagreement. From the fact that an agent is *not* disposed to reconsider, it does not follow that he is not expressing a belief and this is so because even if we accepted a factual sentence

on behalf of others, a disposition to not reconsider could be intelligible and, importantly, could be a sensible response to the type of sentences we accept in public contexts.

In sum, if premise (1) were false, premise (3) could not be true as it stands either because of the fact that the intuition defended by its conditional claim is entailed by both premise (1) and (2). But once the conditional claim expressed by premise (3) is rejected because of the falsity of premise (1), we have to assume that premise (3) cannot be used to test the type of dispositions we exemplify in cases of moral disagreement by taking for granted a disposition rigidly exemplified in factual disagreement. And once assumed that premise (3) cannot be used to support this sort of test because no such fixed disposition is exemplified in cases of factual disagreement, the whole argument collapses.

5. Conclusions

In this chapter I have presented a possible program to solve the assumed *impasse* in *Meta-ethics*. This could be achieved by focusing our attention on second-order issues related to our normative reasons and the way in which the property they express could supervene on certain psychological capacities or dispositions. Accordingly, I have delineated a general argument supporting the general appeal of *Meta-rationality*, which assumes a sound conceptual link between the type of demands posited by our moral requirements and the type of normativity associated with reasons and rationality requirements. From this argument in support of *Meta-rationality* we could offer some insights about the psychological nature of our moral opinions by asking for the structure underlying our uses of concepts such as *reasons for* or *rationality*.

After presenting this general approach, I have proposed three different routes to do *Meta-ethics* that could give content to the conceptual path suggested by the general label of *Meta-rationality*. As Gibbard's program would deserve a treatment in more detail, I have been mainly focused on the two remaining. These options have been introduced by means of two arguments, both of which were aimed to highlight, in a direct fashion, the psychological

nature of our moral opinions by asking for the psychological processes that underlie those cases where our basic normative concepts (*a reason for* and *rationality*) are assured.

Finally, I have defended that even when the arguments had a knockdown-style, the reality is that both are quite problematic as they stand. Moreover, even if the criticisms I have proposed on these arguments were solved, at the end of the day neither of them could offer a solid stance to solve the impasse surrounding *Meta-ethics*.

Chapter 7

Chapter 8

An alternative proposal for *Meta-rationality*

1. The basic idea: *Responding to Reasons*

1.1. Normative facts as facts about reasons

1.2 *Epistemic* and *moral* reasons

2. Do we always have reasons to be rational?

2.1 General remarks about *rationality*

2.2 The normative question about rationality and *the bootstrapping problem*

2.3 A solution for the *bootstrapping problem:* Wide-scope accounts of rationality

2.4 Kolodny's argument against Wide-scope accounts of rationality

3. Defending the basic idea: Reasoning trough second-order normative beliefs

4. Do we have finally have an argument from *Meta-rationality*?

4.1 A Core Assumption

4. 2 Applying the Core Assumption at the level of *Meta-rationality*

4.3. Reasoning, Belief, and the Phenomenology of Necessity

5. Conclusions

Chapter 8

The argumentative structure in last chapter established a negative stance toward the possibilities of *Meta-rationality* as an indirect approach to *Meta-ethics*:

(1)   Current *Meta-ethics* is strongly concerned with the psychological nature of our moral opinions and *rationalism* is a well-established conceptual truth. *Hence, Meta-rationality* could deliver a methodological stance to overcome the *impasse* surrounding contemporary *Meta-ethics* only if we could find a *knockdown argument* highlighting, in a straightforward way, the psychological nature of our judgments of reasons and rationality.

(2)   The *knockdown arguments* that are currently available to highlight, from a second-order stance, the psychological nature of our judgments of reasons and rationality are not as straightforward as we had thought (at least once they are completely unfolded).

        Therefore, *Meta-rationality* cannot deliver a methodological stance to overcome the impasse surrounding contemporary *Meta-ethics* - even when we assume that *rationalism* is a well-secured conceptual thesis

At this point we have a simple choice to make. We can abandon the overall perspective offered by *Meta-rationality* because its best arguments do not offer conclusive answers to restart *Meta-ethics*. Or we can assume that even if these arguments are not well entrenched to restart *Meta-ethics*, yet an additional methodological route from *Meta-rationality* could be depicted to solve the *impasse* surrounding *Meta-ethics*. The basic aim of this chapter is to support the second option. I will focus on a simple question: How should we argue for a methodological shift for *Meta-ethics* based on second-order issues about reasons and rationality once we have accepted that some of the best available routes from *Meta-rationality* are not well founded? As it will become clear at the end of this chapter, I am going to focus on drawing the shape of a general standpoint to do *Meta-ethics* rather than trying to unfold a precise meta-ethical proposal departing from *Meta-rationality*.

1. The basic idea: *responding to reasons*

Until now I have assumed the following principle:

> **(M)** If we want to resolve a set of *meta-ethical* questions (focused, for
> example, on the psychological nature of the state expressed by us when we
> accept a moral sentence), then we must be concerned with some second-
> order questions about our reasons and rationality.

M is a methodological principle. In chapter 7 I suggested that we could remain committed to **M** either by asking some second-order questions about the status of our decisions (Gibbard), by focusing on some contexts where we ascribe a basic normative concept to a certain system (Smith) or, finally, by analyzing cases of normative disagreement (Kalderon). Although attractive, I have argued that these routes fall under criticisms that make impossible to give content to **M** in the way depicted by these routes.

In the current chapter I am going to introduce an additional option to interpret the consequent of **M**. I will defend that if we want to reinitiate *Meta-ethics* from *Meta-rationality* we should start by asking which sort of psychological processes we instantiate when we respond to certain obligations of *rationality*. Thus, I will assume that when we behave *morally* we are responding to certain reasons (*moral rationalism*) that share central features with the reasons we respond to when we are rational, In terms of this basic claim of *parity* between *morality* and *rationality:*

> **(P)** In cases where we *ought* to ϕ because ϕ is *rationally* required by the code
> of rationality and in cases where we *ought* to φ because φ is *morally* required
> by our rules of morality, we are responding to a common type of normative
> fact about the reasons in support of ϕ-ing and φ-ing

Because of (P), we are entitled to give content to the consequent expressed by M in the following way:

> **(RR)** By looking into the psychological processes underlying our responses to what we *ought* to believe or intend in terms of a certain code (*the code of rationality*), we could highlight the psychological processes underlying our responses to the *ought* to do in terms of a certain code (*the code of morality*).

From my point of view, RR offers a plausible piece of advice to overcome the *impasse* surrounding *Meta-ethics* only if the *normative force* of morality can be modelled along the same lines than the *normative force* of rationality, i.e. only if the type of normative fact that we respond to when we are behaving morally is similar to the normative fact that we respond to when we are behaving rationally. This is the assumption in P. Therefore, to sustain the alternative reading of **M** that RR sketches, we should first explain in what sense normative facts about rationality are similar to normative facts about morality.

Accordingly, I will start by noting some similarities between epistemic and moral reasons (section 1.2.). After that (section 2), I will focus on a concern related to the basic assumption I made when I introduced P and RR (section 2.1). I will ask (section 2.2.): What if the normativity of basic requirements of rationality cannot be reduced to the normativity of reasons? What if, after all, we do not always have reasons to be rational? Answering these meta-normative questions about rationality is an essential step to sustain the plausibility of M. But before the importance of these questions can be grasped, I need to sketch the general meta-normative framework in which morality and rationality are related to reasons.

1.1. Normative facts as facts about reasons

Let us suppose that

(1) A ought to ϕ

(1) expresses an *ought fact*. In any *ought fact* a basic normative property - *ought* - is predicated of a certain item (belief, intention, action, desire, etc.) in relation with an agent (A). In my sense, *ought-facts* are *minimal* facts. As far as to assert *that it is a fact that p* is equivalent to assert *that p* (endorsing p only when certain conditions are met), to assert *that it is a fact that A ought to ϕ* is no more than to assert *that A ought to ϕ* in those cases where we

are entitled to do so [Broome, J. 2004. p. 34. See Raz, J. 1999. p. 17-18 for the minimal sense of fact assumed here].

The thesis I want to support in what follows is about the *force* or *normativity* associated with utterances whose logical form is similar to (1). I want to argue, in a nutshell, that the *normativity* of any token of (1) can be explained by referring to a more basic normative relation, which alludes to the following structure

(2) N favours ϕ of A

(2) expresses a *reason fact*[116]. In terms of (2), a certain feature (N) *favours* the formation of a certain attitude or action by A.

The normative relation referred by (2) is usually expressed by means of a reason-predicate. So, those sentences containing a reason-predicate, such as

(3) N is a reason for A to ϕ

 are normally signalling a normative fact similar to (2). The *favouring relation* contained in (2) could *explain* and *justify* the peremptory character of (1) by appealing to a sentence similar to (3). While (1) is a normative fact in a *derivative* sense, (2) is giving voice to an *intrinsically* normative fact [Raz, J. 1999. p. 28. 2001b and 2005. Scanlon, T. 1998. p. 19. Piller, C. 2001. p. 197. Dancy, J. 2001. p. viii. 2004. p. 15 and 29. 2006. note 14. Gibbard, A. 1990. p. 160 and 2003. p. 8. Wallace, R. J. 2004/2006. p. 75. Cuneo, T.2007. p. 65 and Skorupski, J. (forthcoming)].

Four preliminary points need to be stressed about the general image I have just assumed.

The *first point* is about the precise sense in which (3) explains (1) by taking to the foreground a relation like (2). In essence, I am *not* assuming that (3) is *definitionally* basic with respect to (1). A sentence like (3) does *not* unfold the *meaning* of (1). Otherwise: I do not think that (1)-type sentences are expressively redundant because their meaning can be

---

[116] (2) is a relational fact. N is equivalent to what Dancy has recently called a *metafact*, or "(…) facts about some matter of fact and about its making a difference to how to act (…) It is in these metafacts I think of as the central normative facts, to reference to which the normativity of all others is to be explained". Dancy, J. 2006. p. 137. In the same line, Horgan and Timmons write: "Consider the moral reason judgment, 'You should refuse to give the terminally ill patient the huge dose of sleeping medication he requested, because doing so would kill him'. This judgment has the overall form 'Ought p because q', where q is some non-morally described reason" Horgan, T. Timmons, M. 2006. p. 236

completely unfolded by (3)-type sentences. Instead, when I say that the *favouring* relation could explain the force of our basic *ought*, I am simply assuming that we could understand the type of guidance offered by (1)-type sentences in terms of the guidance that (3)-type sentences provide. It is another question whether we would facilitate the same information by uttering a reason-sentence than by uttering an ought-sentence. My basic concern here is about the *normativity* or *force* of certain utterances, not about their *meaning*. Assuming that the force of rationality or morality is not fully grasped in motivational terms, what kind of authority do we presuppose to the code of morality and the code of rationality? [Raz, J. 1999. p. 29-30 Dancy, J. (forthcoming) and Skorupski, J. (forthcoming)].

The *second point* is more controversial. It deals with how the relational nature of the favouring relation could affect the general intuition I have sketched above (that reason-facts could *explain* and *justify* the force or normativity of ought-facts). In a certain way, it is easily assumed that *reasons* – understood, again, as features that favour a certain action or attitude – sometimes interact among them to determine what we ought to believe, feel, or intend. Hence, if in my target scenario A ought to φ because there was a consideration speaking in favour of φ-ing, it is easy to think of a situation where a pair of considerations - one favouring φ and other favouring γ - should be taken into account by A to normatively inform her choice. We tend to claim that in such common scenario the thing to do for A *overall* is determined by the interaction of two *contributory reasons*. Weighing these reasons, we argue, will determine the *overall* duty of A in such situation [Broome, J. 2000. p. 91. Piller, C. 2001. p 196]

Let suppose that in the revised scenario A ought to γ *overall*. If so, we could justify why A ought to γ by appealing to a different favourer than N – N'. Nevertheless, an interesting question emerges at this stage. Once we have determined the duty of A in the new scenario, what happens to the previous favourer – N? If the meta-normative framework sketched above were on the right track - and if N were still present as a favourer- not being completely *silenced* by N' [Dancy, J. 2004. p. 15] – should N help us to *explain* and *justify* the new duty of A? Surely this cannot be true. N, although clearly relevant to *explain* why A ought to γ in the revised scenario, does not *justify* the duty of A to γ. Such duty is grounded on N'. But if this were the case (if N were only relevant to *explain* why A ought to γ in virtue of N') at least a thing would be pretty clear. It is simply not true that in *any* possible scenario we are entitled to appeal to a favourer to *explain* and *justify* an ought-fact. Favourers are cited *sometimes* only to *explain* the occurrence of ought-facts, not to help to *justify* ought-facts in every occasion.

Keeping in mind the revised scenario, a proviso must be assumed about the scope of the general thesis introduced in this section. If I am right, such thesis must be applied with some care. In particular, in those scenarios where two or more considerations constitute our overall duty, the thesis is not true at all. In such scenarios, a favourer is not explaining and justifying an ought-fact. We can cite the favourer to explain why other favourer is requiring something of A, but the explanatory favourer is not justifying by itself the duty of A. Thus, I will focus hereafter on those cases where A is under a *conclusive* or a *decisive* obligation to φ, i.e. cases where the demand expressed by (1) is grounded on a *conclusive* reason for A to φ[117]. As we are basically concerned with the psychological processes that we exemplify in responding to an ought-fact framed in terms of reasons, (1) and (2) above will be understood as referring to *conclusive obligations, obligations* grounded on *conclusive reasons*.

Having said that, the third thing I wish to note is related to the very status of the *favouring relation*. About this particular point, some people defend that the basic normative character we presuppose to *the favouring relation* with respect to *the ought-relation* means that the favouring relation is *not* itself definable in terms of other, more basic, normative relations. Surely this is what Scanlon had in mind when he wrote that '(...) Any attempt to explain what is to be a reason for something seems to me to lead back to the same idea: a consideration that

---

[117] A *conclusive reason* is a slippery notion. In Raz's hands, p is a *conclusive reason* to φ if and only if p is a reason for A to φ and there is no such q that overrides p. See Raz, J. 1999. p. 27. In some other places, Raz accepts that a conclusive reason implies a non-relativized ought-statement. He writes, for instance: "A non-relativized ought-statement states that there are reasons for the fact that are not defeated (…) non-relativized ought-statements mean roughly the same as statements that there is a reason which is defeated only by weighty circumstances or more likely (…) that they are generally undefeated". Raz, J. 1978. p. 13-14.

Dancy, on the contrary, denies that conclusive reasons must be defined by conflating them with what there is most reason to do. *Conclusive* reasons are not *all things considered reasons* of a weighty class. In Dancy's hands, a conclusive reason is "(…) a consideration which, though may be others on both sides, really is the one that decides the issue in the present case. The sum total of the reasons in favour is not itself properly called conclusive; there may be overall more reasons to do the action than not, without any of the contributing reasons being conclusive, and without the combination of them being conclusive either". Dancy, J. 2004. p. 94-95, see also at page 96. Surely, Dancy's sense of conclusive is against a claim recently endorsed by Wedgwood: "A statement of the form 'A ought to φ' seems equivalent to a certain interpretation of the corresponding statement 'There is a conclusive reason for A to φ' (…) it means that A ought to φ all things considered" Wedgwood, R. 2007. p. 24

Although I am not sure if there is a substantive difference lurking here, the point I have in mind is somehow related to the accounts offered by both, Raz and Dancy. As we will see later, reasons of rationality - the ones I will locate later as supporting deontic epistemic facts - are conclusive or decisive in Dancy's favoured sense, i.e. they decide the issue when they are present, without requiring the role of other reasons, and moreover, no matter which other reasons can be identified – although see Dancy, J. (forthcoming) page 5. Equally, and because of that decisiveness, conclusive reasons support, as Raz argues, a non-relativized ought-statement. If I am right, then, our understanding of the relation between (1) and (2) should be narrowed in such way that we do not need to refer to other features than N – where N stands for a fact or for a set of mental states - to explain and justify why A ought to φ.

counts in favour of it' [Scanlon, T. 1998. p. 17]. However, it would be plausible to defend that the *fact* that a certain feature (N) *favours* a certain response (φ) of an agent (A) could be explained by virtue of how certain *values* could be actualized, promoted, respected, or engaged by A's φ-ing in presence of N[118]. Without defending this intuition by now, I want simply to note that the latter claim will be my favoured choice to understand the metaphysical, epistemic and conceptual status of the favouring relation. Surely, the mundane fact that Cesc Fábregas is playing tonight is a reason for you to watch TV because doing so would be your best chance to engage, as Raz sometimes says, with a certain kind of value[119]. If this is so - and insofar as we do not want to leave unexplained how certain natural features are becoming reasons for us[120]- a good option is to appeal to certain *values* to explain the status of the favouring relation [Raz, J. 2001. p. 1. Heuer, H. 2004. p. 133 and 138 and Dancy, J. 2004. p. 33].

Finally, I would like to mention two points about the items that conform the favouring relation sketched by (2). First, let me note a thing about the items playing the N-role in (2). If I am right, N may stand either as an objective fact, i.e. a fact fully independent of A's attitudes (*that she is in trouble, that the building is on fire,* etc.), or as a fact related with A's attitudes (*that A believes that p and that if p, then q, that A believes that she is in trouble,* etc.). If the latter is true, N would conforms a *structural* reason. On the contrary, if the former

---

[118] Raz writes "At the other extreme is the view that all conflicting reasons are merely installations of one value (as in many versions of utilitarianism) and thus fully comparable. But there are a variety of views taking various middle positions, namely regarding reasons as generated by different sources, different and independent values and ideals (…)". Raz, J. 1978. p. 15.

[119] At this point I am assuming a non-humean theory of value. But my general argument in this chapter is largely independent of whether you are sympathetic with a humean account of goodness or not. By humean I understand here someone who will explain why N favours φ-ing of A by saying something like 'in presence of N, doing φ would be desirable for A if A's desires were maximally informed and coherent'. Or alternatively: 'doing φ in presence of N could be advisable (desirable) by A's rational counterpart - A+ - if A+ were located in A's current position'. In both readings, the explanatory status of a favourer (the value the favourer is signalling, so to speak) is dependent on facts about agent's desires. See Railton, P. 1986/2004. p. 10-11 and Gibbard, A. 1990. p. 160. Gibbard writes there: "A fact constitutes a reason for a person if and only if it bears the right relation to his own intrinsic preferences"

[120] Gibbard, for instance, has alluded to this explanatory burden in relation with any non-naturalistic theory of reasons. In Gibbard's sense, non-naturalists (he sometimes cites Scanlon) are committed to assume that certain natural features are reasons without offering any further explanation in support of this brute (normative) fact. To expressivists, by contrary, certain features are reasons because of the fact that these natural features would help to coordinate our plans in a social context. We could track others' plans by tracking certain non-natural features, certain reasons. Because of this simple fact, reasons are sometimes cited as favourers. Gibbard, A. 2007/2002. p. 76. Compare with Gibbard, A. 1990. p. 163. We could locate Raz's position in between non-naturalism in this thin explanatory sense and the just noted non-cognitive story. As I understand Raz on this precise point, a natural feature F is a favourer because a certain response in presence of F would exemplify a value - among a variety of values.

description is correct, N would give voice to a *substantive* or *objective* reason [Pettit, P. 2007. p. 237-238].

Secondly, let me say something about the states favoured by N, i.e. about 'ϕ' in (2) above. Until now I have presupposed that if N is favouring ϕ of A, ϕ is required of A and 'ϕ' stands for a single state. If she is in need, for instance, I have assumed that A ought to help her. Equally, if the building is on fire and A is inside the building, I would say that A ought to run. However, things are murkier than that. Although I will unfold this precise point later in more detail, it should be clear from now that 'ϕ' not always stands for a single state. Think about A's beliefs that p, and that if p, then q. Shall we say that A ought to believe that q? Otherwise: Shall we assume that 'ϕ' stands for a single state? Maybe not. After all, q could be absurd. But if it does not stand for a single state, for what does 'ϕ' stand here? Moreover, if it does not stand for a single state, how does the favouring relation work here?

Ignoring by now questions related with the status of the favouring relation and the precise structure of our requirements (more about the second issue later), let us convene that *ought facts* can be *explained* by appealing to *reasons facts*[121]. We can explain why A ought to help B by referring to certain features in B's situation (the poverty of B, the lack of capacities of B to manage on his own, etc.) [Dancy, J. 2001. p. vii]. These features *favour* a certain attitude of A (in this case the formation of an intention to help B) and explain, as far as that goes, *why* A ought to help B. The fact that A would exemplify certain values by helping B (*generosity*, *respect*, *diligence*, *sympathy*, etc.) could explain why A ought to help B in presence of the just noted features in B's situation. In the same vein, we could explain why A ought to form a certain belief by referring to certain features in her situation, as when we say that A ought to believe that *q* because A believes that *p* and A believes that if *p* then *q*. These non-normative features (A's own beliefs) favour a certain attitude of A (his belief that *q*) and, at the same time, explain why A ought to believe that *q*. The fact that A would exemplify certain values (*coherence*, *informedness*, *cogency*, etc.) by believing that *q* in presence of the just noted non-normative features could explain why A ought to believe that *q*.

---

[121] There are three options available: First, we could assume that the normativity of *ought-facts* could be explained by appealing to *reason-facts* (Dancy, Raz). Second, we could defend that the normativity of ought-facts could not be always explained by reference to facts about reasons, being somehow basic (Broome). And finally, we could assume that neither (i) nor (ii) are accurate to map the shape of our normative domain. We could assume, in line with this latter option, that our normative domain is holistically structured, not supporting (i) nor (ii) (Dancy). See especially Dancy, J. 2006. p. 138 and contrast with Dancy, J. (forthcoming).

But importantly, besides of explaining *ought-facts, reason-facts* can also ground the *normative force* expressed by a significant number of *ought facts*. If I am right, some ought-facts can be *explained* and *justified* by appealing to a normative relation construed around the concept of *favouring*. The fact that you ought to follow a certain rule of etiquette at a given moment could be explained, for instance, by reference to the general structure detailed above. To do so we might specify a set of non-normative facts (which would make reference to your *contingent* membership to a certain group) and a set of *values* that could be respected in case you responded to the previous facts in the required way. Once both factors are isolated, we could explain (a) *why* you ought to follow certain rule of etiquette (you ought to follow it because certain facts are *favouring* your following) and (b) *where* from comes the *force* sustaining your perceived obligation (you ought to follow it because certain *values - group identity, coordination -* would be respected or maximized by your following the rule and such values are related to your current goals).

1.2 Epistemic and moral facts

I have established a general meta-normative framework that stresses the centrality of the concept of *reason* to explain and justify a certain type of deontic normative facts. Once we have this framework in mind, it should be easy to understand in what sense moral facts are similar to epistemic facts, i.e. to facts about the rationality of certain attitudes. Let us consider, for instance, the following utterances

(3) You ought to help B

and

(4) You ought to believe that *q*

(3) and (4) refer to deontic normative facts in the sense depicted in the previous section. While (3) refers to a deontic *moral* fact, (4) refers to a deontic *epistemic* fact[122]. As any

---

[122] Normative facts, in my sense, can be divided in two basic categories. On the one hand we have *deontic* normative facts. (3) and (4), for instance, give voice to this type of normative facts. They demand, as result of a certain *ground*, a certain response (an action or an attitude) from us. On the other hand, we have *evaluative*

deontic fact, their normative force can be unfolded by appealing to the following, more basic, normative relation

(2) N favours ϕ of you[123]

(where ϕ stands for the required attitude, *helping* or *believing,* in these cases). In my opinion, (3) and (4) are structurally similar. They rest on the same type of normative relation, which is expressed by (2). This basic relation, besides explaining and justifying the force of these normative facts, could also help to explain why (3) refers to a *moral* fact while (4) refers to an *epistemic* fact. In terms of (2), the non-normative features that explain and justify the force of (3) differ from the non-normative features that explain and justify the force of (4). While the former refer to the suffering of a certain agent (being moral in the minimal sense of being interpersonally based), the latter refer to a set of mental states of A (being epistemic in the sense of referring to an agent's psychological states) [Cuneo, T. 2007. p. 69-70].

But now consider the following sentence, uttered by a speaker discussing with another person about the appropriate attire to wear for an academic event

(5) You ought to wear a tie

(5) is a normative fact about *etiquette*. As such, (5) expresses an instance of a generic *ought fact*. As any ought fact, the normative force of (5) could be explained and justified by appealing to schemas in line with (2). These schemas would also say why the utterance of (5) conforms a normative fact about *etiquette*. Roughly, it is because of the fact that the non-normative features that explain and justify the force of (5) rest on certain desires of you to belong to a group (*a certain institution*) that (5) expresses a normative fact for you. If I am right then (3), (4), and (5) are normative facts that can be explained and justified by appealing to basic facts about reasons. They share, or so I believe, a basic structural similarity that is

---

normative facts, i.e. which are concerned with the quality of the *response* demanded by a deontic normative fact. For the present purposes, I will be concerned with *deontic* facts. I will use *rationality* and *rational* fundamentally in this sense. So, when I say that you ought rationally to ϕ because a certain ground (G) favours ϕ (whether G is composed by truth propositions or by your own beliefs) I am giving voice to a deontic requirement (objective or subjective) grounded on an evaluative fact. See Smith, M. 2005. p. 14-20. Cuneo, T. 2007.p. 64

[123] Cuneo writes: "Reasons, as I shall think of them, are grounds that favour responses of certain kinds of properly situated agents (…) it is the ground themselves that are reasons. I have not said that reasons are the favouring relation or that they are the ground in addition to the favouring relation" Cuneo, T. 2007. p. 65

grounded on how their *force* could be *explained* and *justified* by appeal to *reasons*, i.e. by appeal to a basic phenomenon in terms of which some features favour an attitude in virtue of how the attitude would respect, promote, or honour a certain value. For the present purposes, however, I want to focus on a more specific similarity established between *moral* and *epistemic* facts. This, as I suggested before, would help us to sustain the methodological piece of advice encapsulated in RR.

Once the previous commonality between (3), (4), and (5) is assumed, we need to ask in what other sense are *epistemic* and *moral* facts similar? In what sense do these facts differ from other type of deontic facts that can also explained and justified in terms of reasons? To many people, the most important resemblance between moral and epistemic facts can be established when we notice the *categorical* character of the reasons that explain and justify moral and epistemic deontic facts. In terms of this shared feature, whether a deontic fact (epistemic or moral) can be *justified* by appealing to reasons is entirely independent of whether the action or attitude that they demand can fulfil certain goals or ideals in the agent. To put in another way: Moral and epistemic facts are *categorical* because the reasons cited by (2) to justify the force of (3) and (4) are *sufficient* by themselves to accomplish this task [Shafer-Landau, R. 2004. p. 199-200 and Cuneo, T. 2007 p. 59].

The point can be better grasped, however, if we attend to a basic distinction that Philippa Foot stressed long time ago. She argued that in speaking about any requirement, we should distinguish the conditions governing its *applicability* from the *reason-giving force* of the requirement itself. Although Foot's distinction was originally used in support of a certain view about the authority of morality, its intuitive form could help to highlight what I want to defend here. In Foot's terms, a requirement can be *categorical* either because its *conditions of applicability* over a certain agent (A) are independent of any goal endorsed by A or because its *reason-giving force* is independent from A's ends or aims. In the same vein, a requirement can be *hypothetical* either because its *conditions of applicability* over A are dependent upon A's desires or because its *reason-giving force* is conditional to A's desires [Foot, P. 2002/1972. p. 158-161].

Once Foot's basic distinction is highlighted, it becomes clear that questions about whether a requirement is hypothetical or categorical cannot be answered in a simple way. Sometimes the applicability and the reason-giving force of a requirement depend upon A's goals or ends (rules of games). Other times, on the contrary, the applicability of a requirement on A is

dependent upon A's desires or goals although its reason-giving force is not (instrumental principle) [See especially Railton, P. 2004/1997 and 2004a]. And finally, there are occasions where it is justified to apply a requirement on A even if neither the applicability nor the reason-giving force of such requirement are dependent on A's desires. Thus, to many people the sense of *categorical* stressed when we equate moral and epistemic facts refers to the last scenario among the ones depicted above. *Moral* and *epistemic* (deontic) facts are categorical, they argue, because *neither* their conditions of applicability *nor* their reason-giving force are dependent upon an agent's goals or desires [Cuneo, T. 2007. p. 59][124]

If they are right, it should be easy to grasp the difference between (3) and (4), on one hand, and (5), on the other. While *moral* and *epistemic* deontic facts are *categorical* in the two senses highlighted by Foot, deontic facts about *etiquette* are *categorical* only in the sense that they can be *applied* to a certain agent without considering his current desires. But they are not categorical in the previous sense. They do not justify A's behaviour in terms of reasons with independence of A's desires or goals [See Foot, P. 2002/1972. p. 161 for a different conclusion, one equating normative facts about *morality* and normative facts about *etiquette*].

Epistemic and *moral* facts share, in sum, a general framework to explain their normative force and a more particular story to unfold their authority, i.e. the precise sense in which they demand some response from us. In terms of this general framework, the force linked to *moral* and *epistemic* deontic facts could be explained and justified by appealing to a primitive normative relation modelled around the notion of *favouring*. In terms of the more particular story, *epistemic* and moral *facts* would give voice to *unconditional* requirements, i.e. requirements that could be applied unconditionally to us (along with requirements of *etiquette*), justifying our behaviour with entire independence of our desires or goals (against requirements of *etiquette*).

---

[124] Pettit writes: "While reasons will count as reasons for an agent like one of us in virtue of their connection with rationality in the broad sense, this does not mean that their normative force as reasons consist, as we agents see it, in that connection (…) If I am rational, I will feel the force of the sort of considerations given, seeing it as a reason. It does not follow, however, that I must see the force of the reasons as deriving as from the way in which it connects with my rationality. I would not find a joke funny if I did not have a sense of humour; but it does not follow that to see that joke as funny is to see it as connected with my sense of humour" Pettit, P. 2007. p. 237

2. Do we have reasons to be rational?

If the attitudes that we exhibit when we are rational are formed in response to certain normative facts and these facts are very similar to the other normative facts that we respond to when we behave morally, why do not shift *Meta-ethics* to issues of *Meta-rationality*? Why not be concerned with the type of capacities that we exemplify when we behave rationally and highlight the types of capacities that we exemplify when we behave morally? Let us ask, at a first stage, for the psychological capacities that we exemplify in response to the reasons that support the force of our requirements of rationality. Once we have an account of these capacities, let us extrapolate, at a second stage, these second-order findings about rationality to our second-order concerns about morality. Insofar as the reasons supporting our epistemic requirements are similar to the reasons supporting our moral requirements the extrapolation would be *prima facie* secured.

However, although this methodological intuition is *prima facie* plausible, a basic problem is lurking from the start. All along the previous section I have accepted a substantive intuition to motivate my methodological approach to *Meta-ethics*. In terms of it, the force expressed by *any* deontic requirement could be explained and justified by means of a more basic normative relation, which can be expressed in terms of reasons. But what if this assumption is not true after all? What if the force of *some* deontic requirements – requirements of rationality - cannot be explained by appealing to the force of the *favouring relation*, i.e. by appealing to certain facts about reasons? In the sections to come, I will focus on this basic problem. Because it has been widely discussed in recent times, I should contextualize it before discussing its implications for my methodological choice to restart *Meta-ethics*.

2.1 General remarks about rationality

As John Broome has recently noted, we usually have in mind two basic senses of *rationality* (or *irrationality*) when we say that a certain system is rational (or irrational). On the one hand, *rationality* sometimes refers to a *property* that A may posses, i.e. *the property of being rational*. When rationality is understood along these lines, that R requires of A that ϕ (where R stands for *rationality* and ϕ for a given attitude-type) should be read as saying that,

necessary, if A *has* R then ϕ. So, if rationality requires that A intends to γ if A intends to ϕ and believes that γ is a necessary means for ϕ then (or so the property-sense argues) necessary, if A is rational then he will intend to γ [Broome, J. 2007. p. 8-9].

On the other hand, *rationality* sometimes refers to a *source* of requirements. In this sense, that R requires of A that ϕ, should be understood as making reference to a *code* or system of requirements (C) in reference to which A's behaviour is being assessed. So, if rationality requires that A intends to γ, if A intends to ϕ and A believes that γ is a necessary means for ϕ then (as far as *the source-sense* is concerned) there is a code or system C which contains a certain requirement (R₁) saying that, necessary, A ought to intend to γ if A intends to ϕ and A believes that γ is a necessary means to ϕ [Broome, J. 2007. p. 12-13].

Although there would be probably a further sense by means of which the two previous senses of rationality could be connected[125], I will assume hereafter that when we speak about rationality we are implying its *source-sense*. Thus, we appeal to *rationality* to evaluate certain items (usually character-traits, dispositions and attitudes) assuming a set of rules of rationality. We should keep in mind *five* claims about this sense of *rationality*. These claims are going to be useful for the discussion to come.

First, I will presuppose in what follows that the requirements of rationality have a conditional structure. Because we have certain attitudes, our basic norms of rationality require certain other attitudes from us. Basic requirements of rationality are mainly concerned with the attitudes that an agent ought to form (or to drop) *in virtue of her other attitudes* [Scanlon, T. 2007. p. 84. 1998. p. 25] Surely, the controversial nature of this remark depends on whether we defend that requirements of rationality are additionally concerned with the attitudes that we ought to form (or drop) in response to certain objective facts about our own situation. Although this sense of rationality gives voice to a legitimate intuition (see chapter 6), in the following discussion I will assume that our basic requirements of rationality *could* govern our attitudes with full independence of whether such attitudes are obeying, in fact, additional and more demanding standards – objective demands of rationality posited in terms of *truth* or *goodness*, for example [Kolodny, N. 2005. p. 509]

Second, a point about the content of our norms of rationality should be stressed. In a certain sense, it is almost philosophical commonsense to point that our basic requirements of

---

[125] In a nutshell: at t_1, t_2, … t_n A would have the *property* of rationality only if A obeyed the *code* (R) - as it applies to him in virtue of his attitudes - at t_1, t_2, … t_n.

rationality govern patterns of attitudes (beliefs or intentions) *without any further qualification*. It is said, for example, that if you intend to P and you believe that your Q-ing is a necessary means to your P-ing, then you are under a basic requirement of rationality, which requires a certain attitude from you. And this is so simply because you are exemplifying the just noted pattern of attitudes and you are minimally self-aware of such fact, i.e. of you exemplifying a certain pattern of attitudes [Broome, J. 2002. p. 89]. For some philosophers this commonsensical characterization is too thin however. Norms of rationality, they argue, are also concerned with those conditional relations that exist among attitudes that have an explicit *normative content*. Scanlon notes for instance:

> "Irrationality in the clearest sense occurs when a person's attitudes fail to conform to his or her own judgments: when, for example, a person continues to believe something (...) even though he or she judges there to be good reasons to rejecting it (...) Irrationality in this sense occurs when a person recognizes something like a reason but fails to be affected by it" Scanlon, T. 25

It is assumed, in this sense, that if you believe that you ought to φ or that you have reasons to φ - where φ stands for an intention, a belief, or a feeling – you are under a somehow *more* basic requirement of rationality, that is, one that could imply a serious amount of irrationality *by your own lights* in case of not obeying it [Scanlon, T. 1998. p. 23-24 and Kolodny, N. 2005 p. 513 for further examples].

Let me say that by noting a certain duality in the formulation of our norms of requirements of rationality I am not assuming that such duality constitutes a remarkable fact by itself. At the end of the day it could be argued that norms of rationality are bifurcated in two different sub-sets of norms. These different subsets of norms, albeit differently formulated, would still refer to certain patterns of coherence or consistence between our attitudes. But they would regulate these patterns in two different ways, so to speak. From this ecumenical stance, I am tempted to think that such duality would suggest a deeper understanding of the nature of the requirements of rationality. It could be established, so to speak, a division between certain possible *codes* of rationality framed in normative terms and other possible *codes* whose basic domain of application would be the functional patterns or relations established among our

unqualified propositional attitudes [Pettit, P. 2007. p. 237]. In what follows, however, I will presuppose a more ecumenical approach. I will accept that there are two basic types of norms of rationality, depending on whether or not a normative content is accepted as the focus of the propositional attitudes over which the rules of rationality apply. Accordingly, I will be concerned with rationality not only in the *narrow* sense, i.e. the sense assumed by Scanlon. I will also assume that rules of rationality are normally posited in non-normative terms. Although at the end of the day I will opt for unifying our requirements of rationality along a certain dimension, by now it is sufficient to accept that rules of rationality are bifurcated in the sense just noted.

The *third* point I want to stress is that *rationality* requires from you only *states of your mind*, *not bodily actions*. By assuming this proviso, I am departing from some traditional accounts of reasoning that depict *rationality* as a mental faculty exemplified through the effective formation of an attitude. In terms of this view, *rationality* is a faculty we exemplify when we draw a proper conclusion, i.e. a belief or an intention, from a theoretical or practical syllogism [Anscombe, E. M. 2000/1963. p. 60. Railton, P. 2004a. p. 179-180]. The explanation of why should we reject this view is simple. If rationality necessarily required from you a certain *act-type* F, then any failure to do F should be considered as a failure of rationality. But insofar as not every failure to do F should be understood as a failure of rationality (because in some cases an external constraint could prevent you from doing F)[126] then we should assume that rationality imposes its requirements only over mental states (beliefs and intentions) and not over your effective actions - at least if we want to avoid undesirable attributions of irrationality [See Audi, R. 2006. p. 89-92 for a recent overview. For a recent statement see Broome, J. 2002. p. 86 and 2005. p. 323].

The *fourth* feature I want to stress about *rationality* is that, despite of the relevance I am giving here to our mental states in fixing our concept of *rationality*, our *evaluations of rationality* are usually understood along a wider dimension. Accordingly, *evaluations of rationality* can be naturally understood either as referring to the fact that an agent's *mental* states obey certain rules of rationality or to the fact that the *agent* falls under a set of standards by means of which we evaluate whether *he* has been rational in forming a certain attitude. In what follows, I will assume that the evaluations focused on the *rationality of the agent's attitudes* and those concerned with the evaluation of the *rationality of the agent by having*

---

[126] If rationality requires from you to F (where F is an act-type) but I am constraining your F-ing, then it should not be assumed that you are irrational for not F-ing.

these attitudes can be carried on separate grounds, with independent rules, presuppositions, and consequences. In many cases, a negative evaluation of an *agent* by having some attitude could coexist with a positive evaluation, in terms of rationality, of the *attitudes* themselves [Bratman, M. 1987. Chapter 6].

Finally, once *rationality* is understood as referring to a *code* of requirements that govern an agent's mental states, it is implicitly accepted that *Logic* has a prominent role in depicting our requirements of rationality. But why is *Logic* so prominent in our commonsensical understanding of rationality? To answer this question let us assume that *Logic*, understood as a philosophical discipline, is primary concerned with the concept of *logical consequence,* i.e. with the property associated with a certain set of propositions (arguments) in terms of which "the (joint) truth of the premises is *necessarily sufficient* for the truth of the conclusion" [Beall, J.C. Restall, G. 2005. p. 1]. Once the focus on the concept of logical consequence is accepted, the type of relation established between *Logic* and *rationality* is quite transparent. As the core concept of *Logic* (logical consequence) is usually encapsulated by means of basic *rules of inference* (which establish whether relations in terms of implication or entailment can be established among set of propositions), a natural route to connect *Logic* and *rationality* would assume that rules of rationality are *applied* rules of inference, that is, rules governing processes of belief-revision or belief-formation in virtue of relations of logical consequence or implication among the content of our attitudes. From a logical rule of inference as *modus ponens* (a rule that I am going to interpret here as saying that if *p* is true and if *p* then *q* is also true then, necessary, *q* is true) we could derive a rule governing our processes of belief-formation. In terms of this rule, if you believe that *p* and also believe that if *p* then *q*, then you ought to form the belief that *q* - or to drop one of your initials beliefs.

Although it is apparently plausible, the problem faced by this picture is that it mistakenly conflates two different levels [Harman, G. 1986. p. 3. Piller, C. 2001. p. 200]. Thus, even if *rules of inference* are somehow relevant for guiding our reasoning, rules of *inference* are *not* rules o*f reasoning*. And this is so because, as David Christensen has recently pointed out:

> "(…) if you learn by the above logical argument [modus ponens] that your beliefs entails Q, you might well want to revise your current belief that P, or your belief that (P ⊃ Q), rather than adopt your belief that Q (…) Modus ponens provides by itself no reason for preferring, e.g.

> becoming a Q-believer to becoming a P-agnostic. In each case, logic gives you no guidance at all regarding which option for revising your beliefs is preferable (…) Facts about the possible truth-value distributions among the member of a certain set of sentences are not diachronic facts about those sentences " Christiansen, D. 2004. p. 5-6.

In sum, as far as processes of belief-revision are *temporally* extended processes carried out among *beliefs* and *Logic* is mainly focused on *synchronic facts* about the distribution of truth-values among a set of *propositions*, *Logic* could be used to fix, at the very best, certain permissible and forbidden combinations of propositional contents. It could not be used as a guide to revise our beliefs from t_1 to t_2[127].

2.2 The normative question about rationality and *the bootstrapping problem*

I have presented five basic features associated with *rationality*. However, if we pay attention to the criteria of relevance endorsed by those who are currently discussing *rationality* topics from a meta-ethical perspective (as opposed to the approach endorsed by people coming from Cognitive Science, Philosophy of Mind or Philosophy of Action), the paramount feature of *rationality* would remain hidden unless we concentrated on a qualitatively different phenomenon. I am referring, of course, to the *normativity* of rationality. Besides the previous features, norms of rationality guide our behaviour with a distinctive *force* or *prescriptivity*[128]. Once this additional dimension is noted, how could we explain the sense in which a certain

---

[127] In this sense, the principle of inference we are using as an example (modus ponens) will *exclude* a certain *simultaneous* set of attitudes - one containing a belief that p, a belief that if p then q but not a belief that q. See Christensen, D. 2004. p. 7

[128] By *normativity* or *normative force* I will simply understand, by now, the property of certain concepts (whatever the metaphysical realm they are picking) by which an *active response* is required once we fall under the concept's domain of application. In a minimal sense, the activity exemplified by a system in a *specifically normative response* to a norm falls beyond of (i) a simple *functional* responsiveness to certain contextual cues encapsulated by functional norms or regularities that enable the system to get its basic necessities or goals - Scanlon, T. 2007. p. 86-87, (ii) a simple case of *correctness* in obeying a norm or a rule – Broome, J. forthcoming a. p. 1-2, and (iii) a motivational reliability to act in agreement with certain norms. In the discussion to come, the relevant sense of activity associated with a specifically normative response to a norm is analytically connected to the recognition of (or reliable response to) those reasons offered by the norm. Normativity is connected with a certain type of authority, *rational* authority.

norm of rationality guides our behaviour? Where lies the *force*, in sum, of our basic rules of rationality?

As I suggested at the beginning of this chapter, there is a plausible route for this. It explains the *normativity* of *rationality* in terms of a more basic normative relation encapsulated by our concept of *normative reason* [Raz, J. 2001b]. Although plausible, this route faces a basic problem. Let us suppose that rationality is basically a system of rules or norms requiring certain attitudes from you in virtue of other attitudes of yours. If this were true, then it should be assumed that you *ought* to form the attitude rationality requires you to form. Inversely, if rationality forbids you to form a certain attitude in virtue of other attitudes of yours, then it should be accepted that you *ought* not to form the state rationality forbids you to form [Kolodny, N. 2005. p. 512. Broome, J. 2005. p. 325]. In general it should be accepted (where R stands for a particular rule contained into the code of rationality and F refers to an action-type) that:

(O) Necessary, if R requires you to F, then you ought to F

Now, let us suppose that, as I defended before, the concept of *reason* is *normatively* basic in respect to the concept of *ought*. Remember that in my sense a concept is *normatively* basic if its normative force cannot be reduced, in principle, to the force expressed by other basic normative concepts. In this precise sense I am assuming that the concept of *reason* is normatively basic in respect to the concept of *ought*. But if this set of assumptions is plausible, then the sense in which we *ought* to form a certain attitude because it is demanded by rationality can be explained by appealing to the following formula:

Necessary, if F is required from you, then you have a reason to F

The formula above endorses a claim to which I will refer hereafter as *non-reductionism about reasons in terms of normativity* - **(NR)** for short. If NR were plausible, as far as the normativity associated to *any* appearance of *ought* could be explained by appeal to the normativity of the favouring-relation, the normative force linked to our basic requirements of rationality - the peremptory or *prescriptive* character of *rationality* - could be explained by appeal to the *prescriptivity* of *reasons* [Broome, J. 2007a]

By applying NR to those cases where we *ought* to form an attitude because it is required by a rule of rationality – cases falling under O - we could obtain a general schema to explain the *normativity* of *rationality* (the sense by which the ought of rationality is requiring you to F) in terms of the normativity of reasons, i.e. in terms of a relation where a feature (A) favours a certain response (F) of you by virtue of how F would exemplify some kind of value or in virtue of how F would fall under the scope of a modal fact about your attitudes [Broome, J. 2004. p. 29, 32-33. Kolodny, N. 2005. p. 512]. This point could be tentatively expressed along the following lines:

> **(NRO)** Necessary, if R requires you to F, then this fact (that F is required by R) gives you a reason to F

NRO could be understood in two different ways. NRO could be interpreted as saying that:

> **(NRO)₁** Necessary, if R requires you to F, then this fact (that F is required by R) gives you a *sufficient* reason to F

Where that F is required by R gives you a *sufficient* reason for you to F is true only if (i) you ought to F because R requires you to F and (ii), necessary, you ought to F if R requires so [Broome, J. 2004. 34-35. 2005. p. 324].

But NRO could also be understood as assuming that:

> **(NRO)₂** Necessary, if R requires you to F, then this fact (that F is required by R) gives you a *pro-tanto* reason to F

Where NRO ₂ is true only if (i') you ought to F because R requires you to F and (ii') it is *not* necessary true that you ought to F if R requires you to F – because that you ought to F is a normative fact dependent on the absence of other codes (S or T) requiring of you not- to F [Broome, J. 2005. p. 325][129].

---

[129] So, when the requirement derived from R is not weighed against other requirements derived from other sources or codes (S or T) we say that the fact that R requires F is a *reason* for you to F. On the contrary, when the requirement derived from R is located in a context where other requirements are in place (derived, for instance, from morality or etiquette) we say that the fact that F is required by R gives you a *pro-tanto* reason to F – a reason that, although implied in the explanation of certain ought-facts, does not directly determine by itself what you ought to do in this precise case. See Broome, J. 2004. Dancy, J. 2004. Chapter 2.

Both readings of NRO face an important problem, however. Let us suppose that, as consequence of your current attitudes (and irrespective of whether such attitudes are well-grounded) a basic requirement of rationality could be applied to you. How could NRO explain the normative force associated with these cases? Niko Kolodny writes:

> "(…) suppose that I have conclusive reasons to have some attitude. In some sense, I ought to have that attitude; it would be irrational of me not to have it. Now suppose that 'ought' here means 'have reasons'. Then we get the bootstrapping result that if I believe that I have conclusive reasons to have some attitude, then I in fact have reasons to have it. This is absurd" Kolodny, N. 2005. p. 512

Kolodny's point can be clarified as follows. He first accepts a basic requirement of rationality:

(K+) If you believe that you have conclusive reasons to F then you ought to F

After that, he considers how we should explain the normativity of (K+) in terms of (NRO). In Kolodny's sense, we should accept either

NRO$_1$ (K+) If you believe that you have conclusive reasons to F then you have *sufficient* reasons to F

or

NRO$_2$ (K+) If you believe that you have conclusive reasons to F then you have *pro-tanto* reasons to F

But neither NRO$_1$ (K+) nor NRO$_2$ (K+) are plausible at all. When they are understood as meta-normative claims unfolding the type of normativity associated with a basic requirement of rationality labelled as (K+), both schemas - NRO$_1$ (K+) and NRO$_2$ (K+) - fall under the

bootstrapping objection. NRO$_1$ (K+) and NRO$_2$ (K+) cannot explain the way in which K+ is normative by appealing to reasons because you *cannot* put into existence a reason to support or justify F (either a *sufficient* or a *pro-tanto* one) by merely *believing* that F is supported by reasons. In the same way (and keeping in mind what I noted above about the two possible ways to formulate basic rules of rationality) you cannot put into existence a reason to γ-ing just because you form an intention to φ and you believe that by γ-ing you will φ-ing. You original intention to φ does not create a reason *out of nothing* to intend what is demanded in order to φ-ing (γ) [Bratman, M. 1987. p. 27-29. Broome, J. 2002. p. 96].

To put it another way: As far as reasons are *relational facts* (either containing a direct appeal to modal facts about the desirability of a certain attitude or act-type under conditions of full-rationality, or stressing the importance of certain substantive values that could be exemplified by forming a certain attitude in presence of certain psychological states - see section 1.1, note 119 above), neither *normative beliefs* nor *intentions* can put by themselves a normative reason into existence in favour of any attitude or action. If this were the case (and this is precisely what is being stated by NRO$_1$ (K+) and NRO$_2$ (K+)) then our normative beliefs would support, in a sort of self-justifying style, those attitudes on which they are focused. And it would be simply absurd [Broome, J. 2001a and 2001. p. 98. Kolodny, N.2005. 512- 514][130]

But if *bootstrapping* is so serious and if it is directly implied by NRO, it seems that at least in some cases it should be assumed that the normativity of rationality couldn't be understood as the normativity of reasons.

2.3 A solution for the *bootstrapping problem:* Wide-scope accounts of rationality

However direct is, the *bootstrapping* problem for NRO could be solved by appealing to the logical form of our basic requirements of rationality, or so John Broome has claimed [Broome, J. 2001. 2001a and 2004]. Remember that, until now, I have been largely

---

[130] Broome - Broome, J. 2002. p. 96 - sketches another line of rejection of NRO. In terms this additional route:

   (1) If NRO were true, instrumental reasoning would be normative only when we have reasons to endorse the state demanded by the instrumental principle.

   (2) But instrumental reasoning is normative – demands something of us – even when we do not have reasons to endorse the state demanded by the instrumental principle (even when we do not have reasons to intend the end on which the instrumental principle applies – along with your instrumental beliefs)
        Therefore, NRO is false.

presupposing some schemata to represent the structure underlying to basic requirements of rationality. I have been assuming, for example, that you ought to align your attitudes in terms of your normative beliefs, expressing this general rule by means of the following schemata:

[**K**] Necessary, if you believe that you have reasons to F, then you ought to F

I have also been concerned with rules of instrumental rationality saying that

[**IR**] Necessary, if you intend to F and you believe that you G-ing is a necessary means to your F-ing, then (as far as G will not be the case unless you intend to G) you ought to intend to G

And finally, I have been concerned with an epistemic rule requiring a certain closure among beliefs:

[**BC**] Necessary, if you believe that p, and you believe that if p, then q, then (as far it matters to you whether q is true), you ought to believe that q

All these conditional requirements of rationality share a common formal feature: their normative force is located in the consequent of the conditional. All of them could be interpreted as saying that if a certain attitude or set attitudes can be ascribed to you (p) then you ought to form a certain other attitude (q). When basic requirements of rationality are understood along this line, we implicitly accept that such requirements have *narrow-scope*. *Narrow-scope requirements* (NS) have the following logical form:

(**NS**) p → O q

(where NS can be read as saying that if p, then you ought to q'). NS expresses a *material* conditional with the normative force attached to the consequent. As any instance of a material conditional, the central feature of NS is that, when its antecedent is true, the consequent is true. Thus, when *p* is the case we can detach, by *modus ponens*, the proposition that you ought to *q* [Broome, J. 2001a. p. 80. Dancy, J. 2001. p. xi].

Let us suppose, in order to see how this formal feature is connected with the normative force of our basic requirements of rationality, that *p* in NS were understood, in fact, as a

variable ranging over a certain set of beliefs (for example, your belief that p and your belief that if p then q). If so, NS would be formally similar to [BC]. But when [BC] is explicitly stated as formally equivalent to a NS-schema, a normative conclusion about what you ought to believe can be extracted in virtue of your current beliefs:

> [BC _{narrow}] If you believe that p and you believe that if p then q, then you ought to believe that q

> *(Factual premise)* You believe that p and you believe that if p then q

> Therefore, by *modus ponens*, you ought to believe that q

Importantly, the same type of normative *detachment* could be obtained by substituting *p* in NS for anyone of the conditions that play the antecedent-role either in [IR] or in [K]. As consequence of this formal feature, any requirement of rationality, if understood along NS, licenses an inference from NS and *p* to Oq.

But let us suppose that the logical form of our basic requirements of rationality is somehow hidden by our colloquial practices. Let us suppose, so to speak, that even when we tend to express our obligations by saying that 'If p, then you ought to q', the logical form of our basic requirements cannot be correctly expressed by a narrow-based schema. Let us suppose that our basic requirements of rationality are telling us which attitudes we ought to form (or drop) if we have another attitudes. Let us accept, in sum, that the requirements of rationality I introduced before were understood in the following way [Broome, J. 2005. p. 322 and 325]:

> **[K _w]**   Necessary, you ought (to F, if you believe that you have reasons to F)

> **[IR _w]** Necessary, you ought (to intend to G, if you intend to F and you believe that you G-ing is a necessary means to your F-ing)

> **[BC _w]** Necessary, you ought (to believe that q, if you believe that p, and you believe that (if p then q))

All these requirements of rationality share a common feature: their normative force (the one I am expressing here by means of *ought*) governs the entire conditional. When basic requirements of rationality are understood along these lines we say that such requirements have *wide-scope* [Broome, J. 2001a. p. 81 and 2004. p. 29[131]]. Wide-scope requirements (WS) have the following logical form:

(WS) O (p→q)

(where WS can be rendered as saying that 'you ought to q, if p'). WS-requirements share two basic features. The first one could be summarized as the negative counterpart, from a formal standpoint, of the main characteristic we associated with NS-requirements. In terms of it, while NS-requirements can be understood as ones permitting *detachment*, i.e. as ones that are making correct the inference going from *p→q* and *p* to O*q*, the main feature of WS-requirements is that they express their normative force in a *non-detachable* sort of way. From O (*p→q*) and *p* we *cannot* detach the conclusion that O*q* [Broome, J. 2001a. p. 82. 2002. p. 93-94. 2004. p. 29-30. Piller, C. 2001. p. 198].

The second feature of WS-requirements, on the contrary, emerges at the moment we ask for their underlying normative force. A standard route to answer the question about the normative force of any WS-requirements would simply point out that such requirements are normative in virtue of how they ban a certain combination of states that would make false the conditional expressed by WS. In essence, as far as WS necessarily bans a certain pattern of states (¬O (¬q ∧ p)), the truth-conditions of WS could be derived from the negation of such banned pattern of attitudes. Insofar as the negation of ¬O (¬q ∧ p) is O (q ∨ ¬p), WS would be true only if O (q ∨ ¬p) were true. Once accepted that the truth-conditions of O (q ∨ ¬p) are conditional - because O (q ∨ ¬p) can be true either by affirming q (if p) or by affirming ¬p (if ¬q) - we should convene that *any* conditional pattern among the ones just referred would make O (q ∨ ¬p) true. And by extension WS[132]

---

[131] Although I am focusing here on Broome, it should be clear that his basic insights about the logical form of our basic requirements of rationality are not news. A similar distinction can be found in Greenspan, P. 1975 - see Greenspan, P. 2007 for an account of the place that issues about the logical form of requirements of rationality would play in shaping the debate around the authority of morality. Additionally, we can track Broome's basic idea in Darwall, S. 1983. Chapter 1. Railton, P. 2004/1997. Blackburn, S. 1998. Chapter 3 and Dreier, J. 1997.

[132] Kolodny writes: "(…) state-requirements ban patterns of attitudes, such as having both A and B. More than one state will satisfy such ban. The state of not having A satisfies it, as well as the state of not having B. So the

To grasp the normative importance of these formal features, let us suppose that *p* in WS were understood, as we did before, as a variable ranging over a certain pattern of beliefs (for example, your belief that *p* and your belief that if *p* then *q*). If so, WS would be formally similar to the previously noted [BC$_W$].

> **[BC $_w$]** Necessary, you ought (to believe that q, if you believe that p, and you believe that if p then q)

[BC$_W$] would ban a certain pattern of attitudes:

> **[BC $_{w'}$]** Necessary, you ought not (to believe that p and that if p then q and not to believe that q)

By negation of [BC $_{w'}$] we could get a direct formulation of the truth-conditions of [BC $_w$]:

> **[BC $_{w''}$]** Necessary, you ought (to believe that q or not to believe that p, and that if p then q)

But once we accept the movement going from [BC$_W$] to [BC$_{W''}$], the formal features that WS-requirements carry with them have two precise consequences. Firstly, from [BC$_W$] we *cannot* detach the normative conclusion that you ought to believe that q if you believe that p, and you believe that if p then q as - we did with [BC $_{narrow}$].

Second, if [BC$_W$] were applied to you in virtue of your attitudes, you would be under a conditional requirement. If [BC $_w$] were applied to you, then you ought either to believe that q (if you believe that p, and you believe that if o then q) or *not* to believe that p and that if p then q (if you do not believe that q). Thus, when [BC$_{W''}$] is applied to you, you are not required to form or drop any separate attitude but rather to satisfy the whole conditional expressed by [BC$_{W''}$].

Keeping in mind this formal feature makes easy to see why the point about *bootstrapping* can be solved by appealing to WS. Remember that this problem emerged when we assumed that it would be possible to explain the normative force of our basic requirements of rationality in terms of the normative force of reasons. It was assumed, in essence that

---

state of not having A is not rationally required. What is rationally required is (either having A or not having B). In other words, the state-requirement has wide-scope. Kolodny, N. 2007. p. 519.

(NR) Necessary, if you ought to F, then you have a reason (either *sufficient* or *pro-tanto*) to F

Once we consider the division between narrow and wide-scope requirements, a more accurate description of how *bootstrapping* emerges follows. In the case of [BR]:

(1) [BC $_{narrow}$] If you believe that p and you believe that if p then q, then you ought to believe that q

(2) NR

(3) You believe that p and you believe that if p then q

(4) By *modus ponens* on (1) and (3), you ought to believe that q

Therefore, by (2) and (4) you have a reason (either *sufficient* or *pro-tanto*) to believe that q

On the contrary, if we accepted a WS-account of the logical structure of our basic requirements of rationality, *bootstrapping* would not be a problem – even if we assumed that NR is true, Again, in the case of [BR]:

(1') [BC $_w$] Necessary, you ought (to believe that q, if you believe that p, and you believe that if p then q)

(2') NR

(3') You believe that p and you believe that if p then q

(4') If (1'), then you ought either to believe that q (if you believe that p and that if then q) or not to believe that p or that if p then q (if you do not believe that q)

Therefore, by (2') and (4'), you have *reasons* either to believe that q (if you believe that p and that if then q) or not to believe that p or that if p then q (if you do not believe that q)

But as far as the conclusion above does *not* license detachment, [BC] does not imply *bootstrapping*. From the fact that you have reasons to believe *q* if you believe that *p* and that if *p* then *q* it does *not* follow - once a WS reading is accepted - that you have *reasons* to believe that *q*. Insofar as you could comply with the wide-scope reason in this case in the other way around (by not believing that *p* or that *p* if then *q*) you could have also a (conditional) reason not to believe that *p* or not to believe that if *p* then *q* in case you were not believing that *q*. So, even if the ought referred in [BC] were explained appealing to the normativity of reasons, from the fact that you believe that *p* and that if *p* then *q* it does *not* follow that you have reason to believe that *q*. You have reasons either to believe that *q* (if you believe that *p* and that if then *q*) or not to believe that *p* and that if p then q (if you do not believe that *q*) [Kolodny, N. 2007. p. 510][133].

In essence, insofar as our basic requirements of rationality can be modelled as WS-requirements, their normative force could be understood in terms of reasons without implying undesirable *bootstrapping*. Evidently, we could question this solution by asking if our basic requirements of rationality are, in fact, WS-requirements. If they were *not* WS-requirements, then - assuming our commitment with NR - the normativity of rationality could not always be explained without avoiding *bootstrapping*. But how could we establish that our basic requirements of rationality cannot be modelled around a WS interpretation? [Kolodny, N. 2005. p. 515 and Broome, J. 2007a. p. 360-361].

2.4 Kolodny's argument against Wide-scope accounts of rationality

Niko Kolodny has recently presented a general argument against WS-requirements [Kolodny, N. 2005]. In order to introduce Kolodny's argument we should note, first and foremost, the importance that certain temporally extended *processes* have to enable us to comply with what

---

[133] Kolodny writes: "If rational requirements have this form, then detachment is not permitted. From the fact that you believe that p, it does not follow that you are rationally required to believe what p entails. So, even if you have conclusive reasons to have the attitudes that you are rationally required to have, it does not follow that you have conclusive reasons to believe what p entails. In sum, understanding rational requirements as having wide-scope leaves us free to understand the normativity of rationality in terms of reasons" Kolodny, N. 2005. p. 515

is required of us by our basic requirements of rationality. In noting the importance of these temporally extended *processes*, Kolodny accepts a general division between two types of requirements:

> "We should distinguish, accordingly, between 'state-requirements', which simply ban states in which one has conflicting attitudes, and 'processes requirements', which say how, going forward, one is to form, retain or revise one's attitudes so to avoid or escape such conflicting states. Any account of rationality that aims to capture our ordinary attributions cannot consist solely of state-requirements, which say 'Not that conflict' It must consist, at least in part, of processes-requirements, which say 'Do this to avoid or resolve that conflict'" Kolodny, N. 2005. p. 517

In Kolodny's terms, both, NS-schemas and WS-schemas can be understood either as expressing *state-requirements* - requirements demanding of you to be in a certain state or combination of states at a given time - or as stating *process-requirements* - requirements saying how to modify your attitudes once a basic requirement of rationality is applied to you.

Let us focus again on the requirement I labelled as [BC]. Understood as *state-requirements*, a NS-schema on [BC] should be read as saying that if you believe at a given moment that p and that if p then q then you ought to believe that q. A NS-schema demands of you to form a *single* mental state - believing q - as consequence of your current mental states (believing that p, believing that if p, then q). On the contrary, interpreted as a *state-requirement*, a WS-schema applied on [BC] would demand of you to believe that q, if you believe that p and that if then q. They do *not* demand of you a single state. Instead, they ban a certain combination of states by taking into account your mental states at a given time. As we saw before, if a WS-requirement demanded something of you, the content of the demand should be put in disjunctive terms, as requiring of you a combination of mental states as consequence of your current mental states - O (q ∨ p).

If basic requirements of rationality purport to be *normative,* however, they should go beyond merely saying that some attitude or pattern of *attitudes* should be formed at a given

time. Understood as *process-requirements* they should also be regulating the range of temporally extended processes by means of which an agent could respond at t_2 to what is required from him by a basic requirement of rationality at t_1. By using Kolodny's own terms: rational requirements must also specify the way in which "one is to form, retain, or revise one's attitude so as to avoid or escape such conflict states" [Kolodny, N. 2005. p 517].

The importance *process-requirements* is evident when we consider how our common evaluations of rationality from a third-person perspective normally work. Let us suppose that at t_1 you were under a basic requirement of rationality understood along the lines offered by WS. Let us agree then that at t_1 you ought (to intend to M, if you believe that you ought to intend to M). Now, let us imagine that at t_2 somebody is evaluating your rationality and he only knows that at t_2 you do not believe that you ought to intend to M - and that at t_1 you believed that you *yourself* ought to intend to M. In a certain way, he must accept - if he is grasping the requirement as a WS-requirement - that you are "as you ought to be". That is, at t_2 he must evaluate positively your rationality in obeying a requirement being applied to you at t_1. Now, let us suppose that in the interval from t_1 to t_2 you have been under the influence of an electrical discharge. As a result of this, at t_2 you do not believe that you ought to intend M. Then, are you rational at t_2? To put it another way: Accepting that at t_2 your mental state is not banned by the requirement applied to you at t_1, could you be evaluated by us as rational at t_2 when such state has been caused by an deviant *process* going from t_1 to t_2?

Kolodny's intuition is that, even if at t_2 you were as you ought to be - at least in terms of the *states* that a WS requirement requires of you - an important dimension of rationality would be hidden from our evaluative standpoint if we did not take into account the nature of the processes going from t_1 to t_2. Thus he concludes that the quality of the *temporal* processes that support obedience to a basic requirement of rationality plays an essential role in our commonsensical evaluations of rationality.

Once this basic distinction between *state* and *process*-requirements is assumed, Kolodny will use it against NS in the following argument [Kolodny, N. 2005. p. 518-520, 527-530. Broome, J. 2007. p. 365]:

(i)   Any genuine requirement of rationality must be normative

(ii)   A particular requirement (W) is normative only if W offered a concrete advice about how to revise the attitudes from t_1 to t_2 to fulfil what is required by W

(iii)   If wide-scope *state* requirements (which demand of you, in virtue of your current attitudes, two conditional states "to be as we ought to be") were *normative*, then they would imply wide-scope *process* by means of which you can *do something over time,* in terms of *reasoning*, to facilitate your fulfilment of what is required by a wide-scope state requirement.

(iv)   *Any* process of reasoning, if rational, should be focused on certain features associated with the *content* of our attitudes.

(v)   By (ii) and (iv), *any* wide-scope process of reasoning aimed to facilitate the fulfilment of what is required by a *particular* wide-scope state requirement (W) should be focused, if rational, on those *content-based* features associated with the states governed by W

(vi)   There are *some* wide-scope processes of reasoning that *cannot* rest on any *content-based* feature of the attitudes governed by W either because (a) there is no attitude at all to ground the reasoning or because (b) even if an attitude were available, it would offers *attitude-related features* to support our reasoning

(vii)   As not every wide-scope state requirement implies a wide-scope process of a rational sort, then - by (ii)-(v) - not every wide-scope state requirement is normative

   Therefore, by (i) and (vii), wide-scope state requirements do not conform genuine requirements of rationality

The argument above establishes that *some* basic requirements of rationality cannot be understood as WS requirements. If such requirements were understood as WS-requirements they would not exemplify plausible processes, temporally extended, of attitude-revision.

   The appeal of Kolodny's argument is in part motivated by the straightforward results of premise (vi). They are directly implied by a certain view of reasoning - sketched by premise (iv) - and by an application of this general view to fix the processes open to us when we fall under a particular requirement of rationality - premise (v). For this reason, I believe that if we were able to reject (iv) and (v), the force of (vi) to support Kolodny's argument against WS-account could be minimized.

In line with this, I will describe Kolodny's case for (iv) and (v). To put my cards on the table, I will try to reject Kolodny's argument by attacking the examples that are adduced to support (vi). In my opinion, WS-requirements are *normative* in the sense defended by (iii) because our conception of reasoning is wider than the limits fixed by (iv). In my view, if (iv) is rejected, neither the type of test proposed by (v) nor the negative consequences stressed by (vi) can be directly derived from (i)-(iii).

Securing (iv): reasoning exemplifies a content-based process

Let me start by (iv). This premise presupposes a general image about our mental states and about the role they play in reasoning. In terms of a well-known approach, *beliefs*, *intentions*, *desires*, *wishes*, and *hopes* are all *propositional attitudes*, i.e. psychological states that can be analyzed in terms of the relation between a propositional content and a certain stance toward it. If I *believe* that the car is broken, the proposition *that the car is broken* is the content of my belief and my belief will be *true* only if the content of such belief is true – only if it is true *that the car is broken*. On the contrary, if I *desire* that the car is broken, although the very same proposition is conforming the content of my attitude (*that the car is broken*), such attitude (my *desiring*) cannot evaluated in terms of truth despite of its shared content with the previous belief. In virtue of the different type of *stance* that I exemplify when I desire something to be the case, my desire could be *fulfilled* only if it *comes to be true* that the car is broken – and not if, at the moment of my desiring, it is true *that the car is broken* [Railton, P. 1994. Velleman, J. 2000. Davidson, D. 2001. Millar, A. 2004].

This basic division between our *stance* toward a proposition and the propositional *content* expressed by such proposition can also help us to find a route to conceptualize the dual role played by certain features in supporting or justifying the processes of reasoning. It is said, for example, that in some cases we ought to form an attitude because we have a reason supporting the *content* of the attitude itself. So, it assumed that we ought to form a belief (B) by taking into account other beliefs of us (C and D) only if the *truth* of C and D would make reasonable for us to believe that B [Millar, A. 2004. p. 42-43. Piller, C. 2006. p. 156][134]. Equally, it is

---

[134] Millar writes: "My believing that it has been freezing may have a rationalizing explanation in terms of my believing that there is frost on the grass and that there would be no frost unless it had been freezing. Here my reason is constituted by the considerations that make up the contents of the beliefs that figure in the *explanans*

sometimes defended that we ought to form an intention to F - where F stands for an act-type - by taking into account the features related to the content of F, which are making good or desirable F-ing [Millar, A. 2004. p. 58-59]. In both contexts, theoretical and practical ones, a reason justifies an attitude in virtue of how it makes explicit some feature related to the attitude's *content,* which justifies the attitude in terms of *truth* (belief) or, in the case of intentions, in terms of the *goodness* linked to certain act-type - F.

Sometimes, however, when we form or drop attitudes by means of reasoning, we are responsive to a different kind of reasons related to the *attitude*-oriented dimension previously noted. Thus, although we have no good reasons in support of the content of a certain attitude (X), the fact that it could be *useful* or *advantageous* for us to form X (to *believe* that X or to *intend* that X) gives us a reason to form an attitude toward X. In all these cases, a valuable feature linked to our forming an attitude favours that we form the *attitude* without being the fact that this feature is supporting or justifying the *content* of the attitude itself. In such cases we say that we have an *attitude-related* reason to X [Parfit, D. 2001. Rabinowicz, W. Ronnow-Rasmussen, T. 2004. Hieronymi, P. 2005. Piller, C. 2006]

Leaving aside whether this distinction is sound, how is it related with (iv)? At a certain stage of his argument, Kolodny mentions an agent that, as result of the psychological states he is exemplifying (he believes both A and B, even if A and B are contradictory), is confronted with a basic requirement of rationality. It demands of him a conditional state to be, by using Broome's memorable phrase, *as he ought to be*. He ought not to believe that A and B at the same time. But how could the agent comply with such requirement? By means of which state of normative *awareness* can he satisfy the conditional requirement that his beliefs impose on him?

Kolodny offers a fundamental insight about how this state of *awareness* should *not* be understood. He writes:

> "It cannot be that the subject reflects on his attitudes
> themselves, recognizes that his attitudes violate a rational
> requirement, and then makes the appropriate adjustments
> on that basis (…) we typically do not comply, and in

(…) my reason is constituted by a consideration that comprise the content of a belief (…) something that I take to be true and to make it reasonable for me to believe that it has been freezing" Millar, A. 2004. p. 12

> some cases cannot comply, with rational requirements in
> this way" Kolodny, N. 2005. p. 520[135]

Kolodny immediately notes - and here comes the relevance of the general remarks noted above – that in order to comply with a basic rule of rationality we should reason from features related to the *content* of the attitudes involved in such WS requirement and not from features related to the *requirement* itself nor to the *attitudes* governed by the requirement - features as *these attitudes could be the rational or beneficial attitudes to have*. He writes:

> "The alternative, I think, is this. From the standpoint of
> attitude A – which has at its object the content of A, not
> attitude A itself – the subject is aware of a need of to
> revise his other attitude, B. Then, on the basis of the
> content of attitude A, the subject revises B. In a broad,
> but recognizable, sense of 'reasoning' the subject reason
> from the content of A to revising B" Kolodny, N. 2005.
> p. 520

And this is what premise (iv) defends[136].

Securing (v) and (vi): applying the reasoning test for wide-scope requirements

Once we accept the general point about the type of reasoning we should instantiate to obey rational basic requirements, premise (v) applies this general lesson to determine whether WS requirements are *normative* at all. To do so, Kolodny formulates a sort of test. He assumes that if WS-*requirements* were normative, they should imply WS-*processes* of reasoning. But

---

[135] See Broome, J. 2006. p. 186-191. Scanlon, T. 2007. p. 85 for two additional rejections of the viability of this particular type of awareness in obeying basic rules of rationality

[136] In Kolodny's account, when we consider what to believe at a certain time, the evidence for the truth of a belief is the paramount normative fact determining what to believe. By saying that we have evidence for the truth of a belief we are simply presupposing that we have some other beliefs supporting - either in a full or in a probabilistic way - the truth expressed by the *propositional content* that is the focus of our doxastic attitude. Richard Moran notes: "With respect to belief, the claim of transparency is that from within the first-person perspective, I treat the question of my belief about P as equivalent to the question about the truth of P. What I think we can see now is that the basis for this equivalence hinges on the role of deliberative considerations about one's attitude (…) And in the case of the attitude of belief, answering the deliberative question is a matter of determining what is true" Moran, R. 2003. p. 62-63

accordingly to the previous remarks, WS-*process* of reasoning, if rational, should enable us to comply with what is conditionally required by a WS-*requirement* by focusing on the *content* of the attitudes involved in the application of the same requirement. Thus, the question asked by premise (v) could be reformulated to grasp the core of what Kolodny has named *the reasoning test* for WS-requirements:

> **[RT]** Could WS-*state* requirements be understood as supporting WS-*processes* of reasoning focused on the *content* of our attitudes?

In Kolodny's terms, if WS-*state* requirements can be modelled in terms of WS-*processes* focused on the *content* of the attitudes involved, WS-*state* requirements are normative - telling us how to change our attitudes from t_1 to t_2 in order to fulfil what is required of us. If, on the contrary, WS-*state* requirements cannot be construed as WS- *processes*, then (insofar as any plausible requirement of rationality should tell us how to revise our attitudes in order to obey it) we should conclude that our basic requirements of rationality, *if normative*, should be understood, by elimination, as NS-based requirements.

Premise (vi) gives voice to the way in which Kolodny resolves this disjunction. In order to illustrate it, I will use Kolodny's example. In applying RT, Kolodny uses a requirement I have previously referred to as K $_{wide}$. I will refer to Kolodny's preferred version of K $_{wide}$ as K hereafter. K expresses, in Kolodny's terms, a general demand to form the attitude (whatever it is) that is supported by the normative beliefs about reasons:

(K) O (p+ →q)

We must read (K) as saying that 'you ought, if you believe that there are reasons supporting p, to believe that p' or, equivalently, as saying that 'you ought, if you believe that there are reasons to intend to F (because F is good along some dimension), to intend to F'[137]. K expresses a wide-scope requirement, demanding a conditional state:

(K') O (q∨¬p)

---

[137] As it is clear by my comment on K, I will use *q* hereafter as referring to any attitude (belief or intention) whose content is the same as the focus of the normative belief expressed by the antecedent of the conditional. Thus, I will refer to an *intention to F* or to a *belief that p* by the propositional variable *q*, using p in all cases to refer to a normative belief about reasons, whose content is either *F* or *p*.

When we apply the *reasoning test* to (K), we get something like the formula below:

(**K'**) O (q∨¬p) →[A *can reason$_C$* from p to q ∧ A *can reason$_C$* from ¬q to ¬p]

K' above must be read as saying that 'if you ought to believe that *p* (if you believe that you have evidence for *p*) or not to believe that you have evidence for *p* (if you do not believe that *p*)[138], then you can reason from the *content* of your belief that you have evidence for *p* to the formation of the belief that *p*, and you can reason from the *content* of your not believing that *p* to a revision of your belief about your evidence for *p*' [Kolodny, N. 2005. p. 527]

Now, take an inverse version of K, in which you are conditionally required not to form an attitude because of the negative evidence offered by a normative belief you have:

(**K***) O (-p→¬q)

K*' should be read as saying something like 'you ought, if you believe that you lack evidence for *p*, not to believe that *p*'. If we want formulate this with respect to practical states, we should interpret it as saying something close to: 'you ought, if you believe that F is not worth pursuing, not to intend F'.

When we apply the *reasoning test* [RT] to the logical factor of (K*) we have:

(**K*'**) O (¬q∨p) →[A *can reason$_C$* from ¬p to ¬q ∧ A *can reason$_C$* from q to p]

(K*') must be read as saying something like 'if you ought not to form a belief that *p* (if you believe that there are no reasons supporting *p*) or to revise your belief about the lack of reasons for *p* (if you *come to believe that p*), then you can reason from the *content* of your belief about your evidence that not-*p* not to form a belief that *p* and you can reason from the *content* of your believing that *p* to a revision of your belief about the lack of evidence for *p*' [Kolodny, N. 2005. p. 528].

Both, K' and K*', are instances of RT. RT grounds the normative character of K and K* upon the availability of two different paths of revision of attitudes in each requirement. These

---

[138] Remember, again, the point just made about the reference of *q*.

two-way patterns are exemplified by the left-hand side of K' and K*'. It is because we are able to revise our attitudes in both directions that K and K* express, in the end, WS-*normative* requirements. The obvious question at this stage, of course, is if such thing is possible. Can we revise, in sum, our attitudes along the patterns offered by K' and K*'?

Premise (vi) offers a negative answer to this question. Let us consider K first. In terms of RT, K should be understood as K'. K' expresses the test imposed by RT upon K. In terms of it:

(**K'**) RR $(q \lor \neg p) \rightarrow$ [A *can reason$_C$* from p to q $\land$ A *can reason$_C$* from $\neg$q to $\neg$p]

Kolodny argues with respect to the possibility associated with this conjunctive pattern of attitudes that, while it is clear that one can reason from the content of the belief from which one has conclusive reasons in support of the very formation of the belief, you cannot reason in a rational way from the content of a non-available attitude - from a belief from which you do not have ($\neg q$) to the revision of your belief about the evidence grounding the attitude ($\neg p$). You do not have any *content* to ground your reasoning in this case. So, if we take K', the fact that there is no content from which we can reason when we consider the second option located in the right hand of K' makes impossible to comply with the requirement expressed by K. A rational requirement like K cannot be rationally resolved in a wide-scope way because we have no content from which to ground our reasoning in the reverse direction[139].

Take K*' now. K*' expresses the test imposed by RT on K'. It claims that

(**K*'**) RR $(\neg q \lor p) \rightarrow$ [A *can reason$_C$* from $\neg$p to $\neg$q $\land$ A *can reason$_C$* from q to p]

In this case, Kolodny would claim that, while we can reason from *the content of one's belief that one's lack evidence for p* to *drop the belief that p* (we can move from $\neg p$ to $\neg q$), it is not clear at all that we can reason from the content of one's believing that *q* to the revision of the one's belief about the lack of evidence for *p*. Although we have now an available content to ground our reverse reasoning, reasoning from *q* to *p* is not *good reasoning* at all. We cannot be *rationally* reasoning from our believing that *q* to a reconsideration of our

---

[139] Kolodny, N. 2005. p. 528

evidence grounding *p*. This is not reasoning, Kolodny argues; it is plain self-deception. So, in K*' we cannot reason in the reverse direction either[140].

Once RT us applied on K and K*, Kolodny concludes that *some* basic requirements of rationality – K and K*, for instance – should be understood, if normative, as NS state requirements. This is so, remember, because in terms of *the reasoning test*, if WS *state* requirements were normative they should exemplify WS processes of attitude-revision. But as far as such processes of attitude-revision are focused, if rational, on certain features associated with the content of the attitudes, WS requirements do *not* imply wide-scope processes. But once we assume that our basic requirements of rationality, *if normative* at all, should be understood as NS state requirements, we are not entitled to affirm that the normative force of these requirements can be reduced to the normative force of reasons. And this is so because affirming such thing would license *bootstrapping*.

3. Defending the basic idea: Reasoning through *second-order normative beliefs*

Let's recap. I am trying to prove that *normative facts about rationality* are explanatory and normatively connected with *normative facts about reasons*. I am trying to secure such meta-normative claim because I think that we could do *Meta-ethics* by asking some questions at the level of *Meta-rationality* (section 1). Linking reasons and rationality, however, faces an immediate pitfall, the *bootstrapping problem*. After introducing the *bootstrapping problem* (section 2.2), I assumed that a possible solution to such problem should move around a WS account of the logical form of our basic requirements of rationality. It was claimed, in essence, that by combining WS and NR we could defend, without incurring in *bootstrapping*, that the normative force of rationality can be explained and justified in virtue of the normative force of reasons (section 2.3). Niko Kolodny, however, has presented the *reasoning test* as a

---

[140] Kolodny writes about (K*'): "There only one direction for one's reasoning to take. This is not because, as was the problem with I+, one has no other content to reason from; the intention to X has a content. It is instead because to reason is to be guided by one's assessment of one's reason (…) The point is not that one shouldn't reason upstream, that progressing upstream is poor reasoning, but that one simply cannot reason upstream, that progressing upstream is not recognizable as reasoning at all" Kolodny, N. 2005. p. 529-530

central case against this solution. This author thinks that once this test is properly unfolded we should convene that *some* basic requirements of rationality are better understood as *narrow-scope* requirements (section 2.4). Particularly, the requirements whose explicit content is formulated in normative terms must be necessary understood as NS-requirements, i.e. as requirements whose logical form is better grasped by means of the formula below:

(NS) Necessary, if you believe that you have reasons to F, then you rationally required to F.

But the problem is that if this were so, i.e. if *some* requirements of rationality should be necessary read as *narrow-scope requirements*, we should accept that any *unqualified* account of the normative force of rationality modelled in terms of reasons incurs in *bootstrapping* [Scanlon, T. 1998. p. 25. Kolodny, N. 2005. p. 540-541 and 2007 p. 3 and 10. Pettit, P. 2007. p. 237. Dancy, J. forthcoming].

Thus, if we put together NR:

(NR) Necessary, if F is required from you, then you have a reason to F

and NS above, we get a very counterintuitive meta-normative claim:

(NS) + **(NR)** Necessary, if you believe that you have reasons to ϕ, then you have reasons to ϕ

Kolodny argues that it is because the counterintuitive nature of this meta-normative claim that we should reject NR as a sensible route to understand the normativity of rationality. However, as far as Kolodny's case against NR is based on the plausibility of a certain account of the process-centered nature of our basic rules of rationality, the obvious question now is whether we should reject NR because of Kolodny's particular view on the processes by means of which we can comply with basic requirements of rationality. Otherwise: Should we reject NR because of the force of *the reasoning test* that he proposes to check the normativity of our basic rules of rationality?

T*he reasoning test* is clearly essential to Kolodny case against NR. Accordingly, we should ask if this test is solidly grounded. In particular, we should ask if Kolodny's version *of the reasoning test* could depict all the variety of reasonable processes of attitude-revision by means of which we comply with the requirements of rationality that apply to us. That is, we

should ask if Kolodny's *transparent* account of attitude-revision is the best way to understand how we are equipped to comply, in a temporally extended fashion, with the requirements of rationality that apply over our attitudes at a given time[141]. In sum, if Kolodny's version of *the reasoning test* were not well-secured (if there were an alternative process of reasoning other than his favoured process to comply with a WS-requirement) then maybe we would find an alternative route to defend a WS-account of rationality against Kolodny's overall rejection.

Kolodny himself acknowledges this possible style of rejection of his view in a recent piece [Kolodny, N. 2007]. There, Kolodny stresses the problems posited *by the reasoning test* on some WS-based processes of attitude revision such as:

(K') RR (q∨¬p) →[A *can reason$_C$* from p to q ∧ A *can reason$_C$* from ¬q to ¬p]

Remember that (K') unfolded *the reasoning test* for the following requirement-type

(K) O (p+ →q)

The problem with (K') is that while it is clear that one can reason from the *content* of the belief from which one has conclusive reasons in support of the very formation of the belief, you cannot reason in a rational way from the *content* of a non-available attitude - from a belief from which you do not have (¬q) to the revision of your belief about the evidence grounding the attitude (¬p).

As I suggested in the previous section, Kolodny assumes that this problem is critical for any unqualified WS account of rationality. However, in the paper I am focusing now he posits a complex question about K'. Why, Kolodny asks, cannot you comply with the this requirement - revising your attitudes in two directions, so to speak - by departing from an attitude *not focused on the content of those attitudes on which the requirement supervenes*? [Kolodny, N. 2007. p. 9].

---

[141] Remember that *the reasoning test* assumes that if WS *state* requirements of rationality were normative, i.e. if they said how to change our attitudes over time to comply with a basic rule of rationality, they should exemplify WS *processes* of attitude revision focused on the *content* of our attitudes. But by focusing exclusively on the centrality of *transparent* features of our attitudes, *the reasoning test* endorses a very restrictive approach to explain how we are capable to resolve the conflict expressed by a basic requirement of rationality in an autonomous way. See Kolodny, N. 2005. p. 520

This is a key question for Kolodny to answer. And it is so because its very intelligibility points toward a possible albeit ignored route to secure a WS account of rationality. In terms of such route, a WS account of rationality could be secured if *other relevant contents* were articulated to enable our compliance with WS requirements of rationality[142]. To put things clear in advance, I think there is at least one type of *contents*, i.e. *other process of reasoning other than Kolodny's favoured one*, that could help us to comply with WS-requirements. Moreover, I believe that by carefully assessing an obvious option we could put in the foreground of the discussion an additional level of *normative awareness* - supplementing those *first-order, content-oriented,* processes of reasoning by means of which we comply with any requirement of rationality. In my sense, *second-order normative beliefs* could help to sustain a WS-based reading of the scope of our basic requirements of rationality[143].

---

[142] Evidently, a cautious note should be made at this point. Kolodny's argument has basically proved that as far as some requirements of rationality must be understood as NS requirements (because they only support NS-processes of attitude-revision) it is *not* necessary true neither that *any* requirement of rationality can be interpreted along a WS schema nor, by extension, that we can explain the normativity of *every* requirement of rationality in terms of the normative force of reasons without incurring in *bootstrapping*. As a consequence, as far as *some* requirements of rationality (those exemplified by Kolodny's cases) would resist an additional rendition in terms of reason (because of *bootstrapping),* the general meta-normative thesis stated by NR is false.

[143] I am going to separate myself from two mistaken attempts to secure a WS process of reasoning, both of which are commented and further rejected by Kolodny.

The first one is focused on the following Kolodny-type requirement

(K*) O (-p→¬q)

Remeber that when we applied *the reasoning test* to (K*) we got the following formula

(K*') O (¬q∨p) →[A *can reason$_C$* from ¬p to ¬q ∧ A *can reason$_C$* from q to p]

In terms of this misguided route, we could revise a normative belief about the lack of reasons supporting a certain attitude F by reasoning from our formation of the attitude not supported by our original normative judgment F. Kolodny will refer to these cases as instances of *upstream reasoning*. In a typical instance of *upstream reasoning* we move from the formation of an attitude (q above) to the revision of the normative belief focused on such attitude (p above) - see Kolodny, N. 2005. p. 534. *Upstream reasoning* could be modelled, Kolodny argues, in two different ways.

Let us suppose that you have formed an attitude - an intention to F - even when you believe that there are no reasons supporting your attitude F. Let us further assume that you were able to form a sort of self-referential belief about your intending F. If both assumptions were plausible, you surely could reason from

(A1) The content of one's belief that one intends to F

to

(A2) a revision of one's belief that he lacks sufficient reasons to F

The problem with this chain of reasoning is quite obvious. Although it can be formally understood as an instance of *upstream reasoning*, it is surely closer to a typical instance of *wishful thinking*.

Now let me introduce the second way to model *upstream reasoning*. In such additional proposal, the intention formed against our normative belief behaves as a sort of *epistemic signal* or *mark*. This *mark* or *signal* would point out the fact that one has sufficient reasons to F by the simple fact that F-ing. Additionally, this *signal* or *mark* would be supplemented by a belief licensing the agent's revision of his normative belief focused on F. By resting on this belief about the epistemic role that his F-ing would have for the assessment of the evidence supporting the content of F, the agent would exemplify a more complex structure of reasoning, one going from

(A1) the content of one's belief that he intends to F

and

(A1') the content of one's belief that if one intends to F, then this would be evidence that one has sufficient reasons to F

to

(A3) a revision of one's belief that he lacks sufficient reasons to F

But the problem with the option above is that once the extension required to secure the rationality of an *upstream processes* of reasoning is accepted - by including (A1') - we cease to be confronted with a *real* or *genuine* instance of *upstream reasoning*, i.e. with a case of reasoning that moves from the formation of an attitude to the revision of a normative belief about the attitude itself. Although the reasoning is located now into a wider framework, this process of reasoning should be better understood as exemplifying a *downstream* case of reasoning, i.e. a reasoning moving from an additional verdict about F to the formation of F - and not as a processes of reasoning moving directly from the formation of F to the revision of the evidence supporting F.

The second proposal I want to separate from, is focused on the other requirement-type Kolodny sometimes focuses on when he attacks a WS account of rationality:

(K) O (p+ →q)

As I noted above, from (K) we get the following formula once we apply the reasoning to it

(K') RR (q∨¬p) →[A *can reason$_C$* from p to q ∧ A *can reason$_C$* from ¬q to ¬p]

Kolodny considers the following proposal about (K'): If our *not* forming a certain attitude in presence of a normative belief supporting such attitude (¬q above) were accompanied by a belief whose content is *that there is insufficient evidence that someone has reasons to p*, could we resolve the conflict regulated by (K) in both directions? Otherwise: Could we satisfy (K) by being able to form the attitude supported by the normative belief and, the other way around, by being able to revise the normative belief - once a certain second-order belief about the reliability of our reasons is assumed?

Kolodny's remarks about this option are straightforward. By assuming that a belief about the reliability of our reasons could help us to resolve rationally the conflict regulated by (K), we are *not* responding to the conflict expressed by (K). In fact, we are responding to a different normative conflict, which is expressed by the following requirement:

B-NS: If one believes that there is insufficient evidence that p, then one is rationally required not to believe that p

B-NS expresses – as result of our forming a second-order belief about the evidence- an additional requirement. Accordingly, the fact that we can drop our normative belief - as required by (K') - does not follow from an *upstream* process of reasoning but rather from the fact that an additional NS requirement (B-NS) is operating on our attitudes [Kolodny, N. 2005. p. 533]

To motivate the role of *second-order normative beliefs* in reasoning I will follow a three-stage route. I will argue first for the plausibility of this type of beliefs by rebutting some critical remarks made by John Broome. After that, I will try to secure the role of second-order normative beliefs by explaining why it would be good for us to be able to form such beliefs. I argue there that they would help us to further secure our capacity to respond to reasons in theoretical or practical reasoning. And finally, I will explain the function of second-order normative beliefs even in those cases where our basic rules of rationality are narrowly formulated. As far as Kolodny appeals to these kinds of rules of rationality, we should be able to motivate the role played by second-order normative beliefs also in this narrower normative context. In what follows I will try to secure a case for second-order normative beliefs by attending to the dimensions just mentioned. Let me start by Broome's recent criticism [Broome, J. 2006 and forthcoming].

First Task: Arguing for the overall plausibility of a second-order model of reasoning

John Broome focuses on the *conscious* processes of *reasoning* by means of which we comply with a basic requirement of rationality. In Broome's view, *reasoning* is "an activity - something we do – through which we can satisfy some requirements in particular instances" [Broome, J. 2006. p. 184]. Thus, although some reasoning is *unconsciously* developed, Broome's discussion moves around the processes of reasoning accessible to certain level of normative awareness, i.e. *conscious* processes of revision of attitudes. Broome is interested in these processes by a simple reason: If Kolodny's central intuition is right, then there must be a process of reasoning through which *we* (understood as agents and not as a set of sub-personal processes or mechanism) could satisfy *any* presumedly requirement of rationality. Thus, given any requirement of rationality (R) we could ask for the processes of reasoning by means of which *we* could comply with it. If we could not isolate any agent-centred process of reasoning that could bring us to comply with R, we should establish that R is not normative at all. In Broome's words:

> " (…) if it should turn out that no process of reasoning
> could bring you to have, say, transitive preferences, that
> would cast some doubt on the claim that rationality
> requires of you to have transitive preferences. We would
> certainly want an explanation of how there could be this
> requirement on you without your being able to bring your
> self to satisfy it " Broome, J. 2006. p. 185

Broome acknowledged that questioning the nature of our processes of reasoning is not a very common practice. To many people, the most exciting problem about our requirements of rationality is not related to the nature of our processes of reasoning but rather to the rules of rationality. How to formulate them, how many there are, and how are they related to each other are, accordingly with a widely assumed approach, the basic questions about our rules of rationality [See Gibbard, A. 1990. Chapter 1].

Surely, the centrality of these substantive concerns could be explained because of the prevalence of a default image about the nature of our reasoning. Broome points out that some people tend to assume that the important questions about rules of rationality are substantive ones because such default perspective is not taken to be problematic. In terms of this view:

> "(…) you can reason your way to satisfy a requirement
> by starting from the requirement itself as premise (…)
> from your believing some proposition such as the ones
> that I have labelled *Modus Ponens, Necessary Means*, or
> *Transitivity*" Broome, J. 2006. p. 185

This *received view* - 'an all purpose model' as Broome calls it [Broome, J. 2006. p. 185] - accepts that the processes of reasoning aimed to bring ourselves to satisfy a rule of rationality set out from a belief in the requirement itself [Broome, J. 2006. 187]. In contrast, Broome defends that this intuitive approach is unstable and that a first-order model of reasoning would be preferable to understand how we comply with our basic requirements of rationality. Hence, I believe that we could identify a good basis for a second-order account of reasoning by carefully considering Broome's criticisms on the received view.

In order to criticise *the received view*, Broome focuses on the epistemic requirement labelled *Modus Ponens* or *Belief Closure* – MP for short. In terms of MP:

(MP) O $_X$ (to believe that *q* if X believes that *p* and X believes that if *p* then *q*)

Let us suppose that your mental states fall under the scope of MP. If *the received view* is right and you could satisfy any requirement through a second-order normative belief, then *you* would be able to satisfy MP starting from the following second-order belief:

MP$_{B ELIEF -WS}$ B $_I$ (O $_I$ (to believe that *q* if I believe that *p* and I believe that if *p* then *q)*

According to Broome, this self-referring normative belief licenses certain pattern of reasoning. It is because you exemplify an MP $_{BELIEF-WS}$ that you can *assert* to *yourself* the following content[144]:

I ought to believe that *q* if I believe that *p* and I believe that if *p* then *q*

And it is because you assert to yourself the previous content that you can *express* to yourself the following *intention* by telling yourself 'so'

I shall believe that *q* if I believe that *p* and I believe that if *p* then *q*

This process, Broome argues, is a piece of reasoning[145]. Through your second-order normative belief you can form an intention that will be fulfilled in due course [Broome, J. 2006. p. 187. See Broome, J. 2002. p. 85]. As my formulation makes clear, it is because the content of MP $_{BELIEF}$ is WS that the intention that you are expressing in this pattern of reasoning is WS [Broome, J. 2006. p. 188].

---

[144] In Broome's terms, the function of these acts of saying to ourselves something can be basically explained in terms of how these acts bring together certain beliefs. It is because we can put together certain beliefs (that I believe that p and that I believe that if p then q) through these types of acts that we can acquire other mental states (the belief that q) as a result of our previous mental states. See Broome, J. 2006. p. 192

[145] Because reasoning, in Broome's sense, is "the causal process whereby some of your mental states cause you to acquire a new mental state" Broome, J. 2006. p. 192. Broome, J. 2002. p. 86. Compare with Harman, G. 1998/1976. p. 149

Broome notes, however, two basic difficulties for *the received view*. The first one is related to the *scope* of your intention. Broome notes that as far as the intention expressed by you is WS, you cannot end your reasoning by specifically believing that *q*. To do so, you would have to start from a NS second-order normative belief. But if our basic requirements of rationality are WS, neither you are departing from a NS second-order normative belief such as:

> **MP** $_{\text{BELIEF - NS}}$ B $_{\text{I}}$ (if I believe that *p* and I believe that if *p* then q, then O $_{\text{I}}$ (believe that *q*))

nor you can *assert* to yourself that:

> I ought to believe that *q*

So, as a matter of fact you cannot *express* to yourself a NS-focused intention in the line of:

> So, I shall believe that *q*

And if you *cannot* even express this intention to yourself you are not going to believe that *q* in due course.

Although direct, this difficulty is not very serious anyway. Defenders of *the received view* can argue, as a sort of rebuttal, that our intentions in these contexts of reasoning can be narrowed without appealing to any normative belief. In doing so, they might note that MP can be applied to you only because you exemplify certain beliefs. It is because these beliefs are exemplified at the moment I am asserting to myself the second-order normative belief that I can narrow a WS-intention. Thus, as far as MP is applied to me only if:

> I believe that *p* and I believe that if *p* then *q*

I can form the very same second-order normative belief that is required to put *the received view* to work:

MP BELIEF-WS B $_I$ (O$_I$ (believe that q if I believe that p and I believe that if p then q

I can *assert* to *myself* the very same content that I noted before:

I ought to believe that q if I believe that p and I believe that if p then q

I can *express* to myself the very same intention:

So, I shall believe that q if I believe that p and I believe that if p then q

And I can end by believing that q because, as a matter of fact, I believe that p and that if p then q. I can form the belief that q, in sum, because one of the routes permitted by the WS requirement I am under is *actualized* by my believing that p and my believing that if p then q.

Broome recognizes that those defending *the received view* can appeal to this route to minimize the first criticism [Broome, J. 2006. p. 189]. However, even if they solved the first problem in this way, a much more demanding objection would be lurking. This objection represents the *second* and most important problem for *the received view*. Leaving aside questions concerned with the scope of our intentions, this criticism assumes that:

> "Intending to believe a proposition is normally ineffective; it normally does not get you to believe the proposition" Broome, J. 2006. p. 189

Remember that *the received view* assumes the mediation of an *intention* to explain how we arrive to the state required by a particular rule of rationality. However, according to an extended orthodoxy, our beliefs cannot be directly channelled by our intentions [Williams, B. 1971. Hieronymi, P. 2006]. Thus, *the received view* (at least when it tries to explain how we comply with requirements of epistemic rationality) is false: Our intentions cannot control our beliefs.

Broome thinks that this objection is decisive for the plausibility of *the received view*. But he recognizes that this is so only in a conditional sort of way. The objection is decisive, so to speak, only because the other possible route to construe *the received view* (the one that does not require the mediation of intentions to channel the influence of our second-order normative beliefs) is not going to work either. In terms of this unsuccessful route, our second-order

normative beliefs could channel any demanded state without requiring the mediation of an intention. If this option were plausible, I could move from a second-order normative belief such as:

MP **BELIEF-WS** B $_I$ (O$_I$ (believe that *q* if I believe that p and I believe that if *p,* then q*)*

to the act of *asserting* to myself a certain (normative) content:

I ought to believe that *q* if I believe that *p* and I believe that if *p* then *q*

And from this act, which overlaps with the utterances of certain expressions such as 'so' and 'then', I could pass directly to believe that *q*.

Broome identifies this failed alternative with a central intuition endorsed by Scanlon [Broome, J. 2006. p. 190]. Remember that Scanlon assumed that our most basic attitudes (intentions and beliefs) could evolve, other things being equal, in response to a variety of second-order normative beliefs – in response a variety of judgments about the reasons supporting certain attitudes. In Scanlon's sense, *judgment-sensitive attitudes* could be formed or revised in response to these types of judgments. Thus, if you believed that there are reasons to ϕ (where ϕ stands for an attitude), you could directly ϕ-ing without previously forming an intention to ϕ [Scanlon, T. M. 1998. p. 20-24 and 67]. Broome notes that this cannot be true. Let me quote the entire passage where he rejects Scanlon's view:

> "If you believe that you ought to have some belief, that would not normally cause you to have the belief. Suppose that you ought believe that you are attractive, because believing that you are attractive would make you more approachable (…) This would not normally cause you to believe that you are attractive. Normally our beliefs are caused by evidence, not by normative beliefs about what we ought to believe (…) You would believe that you are attractive when you judge that there is sufficient evidence that you are attractive. Beliefs are genuinely judgment-sensitive in this sense, but it is not Scanlon's sense. Your

> judgment in this case is about the *content* of the belief,
> not about the *belief* itself. It is a first-order belief, not a
> second-order one" Broome, J. 2006. p. 19. Cursives
> mine.

The core of Broome's criticism moves around a dilemma. Either our (second-order) normative assertions are expressive devices aimed to signal the presence of causally relevant first-order beliefs, or they are supposed to point toward second-order beliefs, beliefs that can cause the demanded state by themselves (without the presence of any first-order belief). If the latter is true, Broome will ask how could we explain that your normative belief about your attractiveness cannot help you to believe that you are attractive? Otherwise, if the former is true, Broome will ask why model our reasoning around the functional role of second-order normative beliefs if the causal effectiveness of such reasoning is sustained, in fact, by our first-order beliefs? In either interpretation, in sum, the role played by Scanlon's normative beliefs is problematic.

Is it possible if faced with this dilemma to remain faithful to *the received view* in Scanlon's sense? Can we argue, in sum, in support of the only route that would be able to secure a place for second-order normative beliefs in our reasoning processes without requiring, at the same time, the intervention of an intention to form the required attitude? To answer these questions, let us assume that if Broome were right, the most plausible way to interpret the role played by *second-order normative beliefs* would accept that when we *assert* to ourselves the following content:

> that I ought to believe that I am attractive

we are giving voice to a set of first-order beliefs that support the truth of *content* expressed by the required belief, i.e. *that I am attractive*. To put it another way: I can believe that I am attractive by taking into account the assertion above because this assertion channels a previous relation of evidential support between some of my first-orders beliefs and the demanded content (*that I am attractive*). It is for this reason that no second-order belief is required to explain how we believe that we are attractive when we believe that we ought believe such a thing. We can move directly from asserting a normative statement to forming the state it requires without the intervention of a second-order belief (and without appealing to

an intention to form the required state). All we need to assume, Broome suggests, is a set of first-order beliefs and a certain story about the expressive capacities linked to self-focussed assertions involving *ought*. Once this is in place, we do not need to appeal to second-order normative beliefs to explain why we end by believing that we are attractive at t_1 when we assert to ourselves that we ought to believe that we are attractive at t_0.

But surely, even if Broome's treatment of this particular example were accurate [see Crisp, R. 2006. p. 37] it is not going to be easy for him to dismiss the *overall* plausibility of *the received view*. At least not without making some important distinctions that come to mind when we consider the different functions that we usually associate with the utterance of a basic normative concept. Let us consider *wrongness,* for instance. Returning to Scanlon, he distinguishes three different senses associated to *wrongness*. First, *wrongness* "might be what I called a *buck-passing notion*, indicating the presence of other reason-providing considerations, rather than a reason-providing notion" [Scanlon, T. 2007a. p. 6]. Second, he points toward a *shaping sense* of *wrongness* by means of which this concept would shape "(…) the way I should think about the decision I face (…) determining which considerations I should take to be reasons but not offering itself a reason to behave in a certain way" [Scanlon, T. 2007a. p. 7]. Finally, *wrongness* is sometimes used as a *backstop* concept, i.e. as a concept providing *itself* a reason to behave in a certain way. According to this latter sense, if X is *wrong* then we have a reason not to X-ing because of the fact that X-ing would be incompatible with governing oneself "in a way that others could not reasonably reject" [Scanlon, T. 2007a. p. 8].

Keeping these different senses in mind, let us suppose that the particular normative assertions picked out by Broome to dismiss the plausibility of Scanlon's version of *the received view* exemplified, in fact, what Scanlon has sometimes referred as a *buck-passing* use of *ought* [Scanlon, T. 1998. p. 95-100. 2007a. p. 6. Crisp, R. 2006. p. 63-67]. According to this, when I assert to myself

That I ought to believe that I am attractive

I am not predicating some normative property of my beliefs. Neither I am expressing any second-order belief (for instance: the belief that my belief has such normative property).

Rather, what I am really making explicit is the availability of a set of first-order beliefs that could support the truth of *p* (*that I am attractive*) and that could cause my believing that *p*[146].

Nevertheless, even if a *buck-passing* account of *ought* is quite central to understand some core normative utterances, a generalized *buck-passing* model cannot explain every possible use of *ought*. In particular, when I assert that

I ought to believe that *q* if I believe that *p* and I believe that if *p* then *q*

I am not requiring other mental states than my belief that *p* and my belief that if *p* then *q* to guide my use of *ought*. Such states guide my assertion, to some extent, without necessary referring to an additional set of first-order beliefs. The *ought* of rationality can be properly applied in these cases (the ones expressed by *MP* and the *instrumental principle*) over a set of attitudes (S) without taking into account the evidential status of S, i.e. the support that other first-order beliefs offer to S. But if this were so, we should convene that the utterance of *ought* is not always governed by a *buck-passing rule*. That is, on some occasions we do not need to assume the availability of a set of first-order beliefs to explain (and later justify) a given utterance of *ought*. The *ought* of rationality, in sum, does not always behave like *a buck-passing* predicate, i.e. like a predicate whose basic role is to refer to other first-order beliefs that evidentially support the required state. But if the *ought* of rationality cannot be always understood as a *buck-passing* term, which additional function exemplifies?

In my view, the utterance of *ought* noted above can be explained by appealing to a *backstop schema*. In terms of this schema, *ought* is not always a *reason-tracking* predicate. Sometimes *ought* is behaving like a *reason-providing term*. When, as in the case above, we assert to ourselves that we *ought* to ϕ - where ϕ stands for a conditional attitude - we are not necessary presupposing the presence of a set of first-reasons that would support any of the demanded states – *explaining* and *justifying* why we ought form them. On the contrary, when we assert *ought* in such occasions, the predicate tracks by itself a *reason* related to the value that

---

[146] Broome writes: "A second-order judgment of this sort often accompanies a first-order one. When you judge there is sufficient evidence for some proposition, you may well to judge that you have sufficient reasons to believe the proposition. But what causes you to believe the proposition, if you do, is the first-order judgment, not the second-order one" Broome, J. 2006. 191.

coherence and *consistency* have for our sense of responsible agency [Pettit, P. 2007. p. 232-233][147].

In terms of a *backstop* account of *ought*, as soon as our attitudes are related by means of certain patterns (usually ones mirroring some basic rules of logic), there is a value that could be respected by forming or dropping a certain attitude and, consequently, there is a reason that would favour either the formation of a new attitude or the revision of the previously formed attitudes. Leaving aside whether the attitudes that enable the application of a given rule of rationality are well supported by themselves, the *backstop* story claims that at the very moment we exemplify a certain pattern of attitudes (that *p* and that if *p* then *q* or that *p* and that *not-p*) we are in a position to respond to a certain reason. It is in virtue of how our forming *any* of the demanded states could constitute an appropriate response to a basic value, the value that *consistency* has for our deep sense of agency, that we can explain and justify our forming or dropping an attitude. Thus, accordingly to a *backstop* account, by uttering:

I ought to believe that *q* if I believe that *p* and I believe that if *p* then *q*

we can express (against Broome's suggestion) a second-order normative belief focused on the following content:

MP **BELIEF-WS** B $_I$ (O$_I$ (believe that *q* if I believe that *p* and I believe that if *p,* then *q)*

This belief responds to a *structural fact* [Scanlon, T. 2007a]. The application of a requirement of rationality *in virtue of our beliefs* could offer by itself a (complex*)* reason to form or to drop a certain attitude or attitudes - leaving aside whether the demanded states

---

[147] *Coherence,* hereafter, refers to a paramount kind of value. *Coherence* – understood, again, as internal consistency in one's mental states – is important because it allows us to pursue things of value. Raz, J. 2005. p-17 as cited by Wallace, R. J. 2001/2006. p. 113. Thus, although it is conceptually possible to imagine a coherent and yet crazy person (Gibbard, A. 1990. p. 157) it is nonetheless conceivable to imagine an agent that is unable to engage with substantive value because of his incapacity to be coherent. In line with this intended sense, Shafer-Landau notes that "we ought to value coherence because we value what coherence makes possible" - Shafer-Landau, R. 2003. p. 28. Equally, Bratman appeals to *coherence* as an essential ingredient of what he denominates *primary goods* in the life of an agent. Bratman writes: "Our capacities for such intention-like attitudes, and structures of planning and practical reasoning, are, for temporally persisting and social agents like us, more or less all purpose, universal means; they are means to an extremely wide range of divergent human goods". Bratman, M. 2004/2007. p. 283

could be justified[148]. Thus, *ought* sometimes track this particular type of reasons - *reasons of consistency*. And it is because of this *backstop* function that ought-utterances can be related to a second-order normative belief, i.e. a belief focused on a *reason of consistency*. Therefore, in terms of a *backstop* account some utterances of *ought* can be associated, at least in the context of certain assessments of rationality, with a second-order normative belief being responsive to the value of *consistency*. In terms of this association, a second-order belief focused on such value accompanies your belief that *p* and your belief that if *p* then *q* when you assert to yourself that

I ought to believe that *q* if I believe that *p* and I believe that if *p* then *q*

Such belief supposedly tracks, to put it some way, a paramount value for your own agency (*consistency*) to which your can respond either by exemplifying a certain attitude or by revising a set of previous attitudes. Although the process of revision is carried out by means of a first-order process of reasoning, one focused on evidence, your ability to form second-order normative beliefs is fully independent of whether your mental states are well-supported by evidence. In a certain way, by forming a second-order normative belief you are responding to a certain reason, a conditional one.

Second Task: Arguing for the utility of second-order normative beliefs

Let us suppose that second-order normative beliefs are not completely implausible. As I noted above, even if we assumed that such beliefs are not completely implausible, we would need to secure them further in two different senses – at least if we want to appeal to them to support a WS-processes of reasoning. Firstly, we need to ask for the *function* that second-order normative beliefs play in our psychological economy. Which utility could be derived from our capacity to form second-order normative beliefs? Which function could be fulfilled by such beliefs in the context of self-governance and self-management? Secondly, even if we had offered a general explanation of the role of second-order normative beliefs in our processes of

---

[148] In my favoured sense, a *second-order normative belief* supervenes on a set of first-order beliefs only if a second-order belief about your own beliefs is in place. In the case at hand, the second-order *normative* belief is supervening on (i) your first-order beliefs that p and that if p then q and (b) your (second-order) belief that you believe that p and that if p then q. Se Shoemaker, S. 1996. p. 29 and 34 for this general point.

reasoning to comply with basic rules of rationality – *consistency, belief closure*, *instrumental principle*, etc - we should offer yet an additional explanation focused on the role of such beliefs in cases in which rules of rationality narrow formulated apply, i.e. rules whose applicability is dependent upon normative beliefs focused on our attitudes.

Philip Pettit has recently highlighted the general importance of second-order normative beliefs for reasoning [Pettit, P. 2006. 2007, and 2007a]. He has offered an explanation focused on the first question noted above. By attending to some of his remarks we can isolate a wider explanation of the functional rationale underlying our capacity to form second-order normative beliefs in the context of reasoning. Additionally, by assuming Pettit's story we could explain how we can comply with rules of rationality narrowly formulated. Thus, by pressing further on the prospects of a general theory of reasoning containing second-order normative beliefs (as the one proposed by Pettit) we could debunk Kolodny's own style of criticism, i.e. one focused on requirements of rationality narrowly understood.

Let me begin with Pettit's general view on reasoning. Pettit starts by assuming that we can comply with certain *constraints* or *desiderata*[149] of rationality in two different ways. He focuses on MP, as I did before. Pettit argues that once we are under the normative guidance of MP because of our beliefs (because we believe that *p* and that if *p* then *q*), we can comply with MP either by directly believing that *q* or by forming a *meta-representation*, i.e. by forming "a belief about the beliefs involved" in the application of this particular instance of MP [Pettit, P. 2007. p. 233]. Pettit notes that by taking the former route:

> "The system holds the belief that if p, then q (…); it
> comes to form the new belief that p (…) and then, more
> or less automatically, he it goes on to form the belief that
> q (…)" Pettit, P. 2007. p. 233

---

[149] In Pettit's sense, a rational agent is an agent that functions more or less properly, i.e. an agent that embraces goals that can be fulfilled, that forms accurate representations of his surroundings and, finally, that "acts for the realization of those goals according to these representations" Pettit, P. 2007. p. 232. *Constraints* or *desiderata* of rationality, in Pettit's sense, are rules that allow the system "to operate satisfactory in its own terms" Pettit, P. 2007. p. 232. Pettit distinguishes two basic types of rules of rationality: On the one hand, he stresses the importance of certain *constraints* of *compossibility* and *consistency*. In Pettit's sense these rules are essential to secure the agent's performances: "Let the agent try to act on an inconsistent set of representations, and it will size up (…) Or let try to enact an inconsistent set of goals (…) and it will also stall" Pettit, P. 2007. p. 232. See also Pettit, P. 2006. p. 1085-1086. On the other hand, Pettit accentuates the importance of certain *desiderata* of *closure*. In terms of these rules, a system must be able to see "what its current representations or goals entail in this way, since this might have an impact on what it does (…)". Pettit, P. 2007. p. 232. Again, the force of these rules of rationality rests on how they sustain our own agency: "Let the system satisfy those desiderata of closure, and it will be better fitted to perform as agent" Pettit. P. 2007. p. 233.

On the contrary, by exemplifying the latter path

> "(…) we will be able, however implicitly, to register that proposition of the form 'if p, then q' and 'p' entail the truth of the conclusion 'q'. We will form a belief in that entailment, a sensitivity to the connection involved. Such belief is a belief about the propositions involved, not a belief in those propositions. It is a higher-order or meta-propositional belief: a belief in the meta-proposition "'p' and 'if p, then q' entail 'q'" Pettit, P. 2007. p. 253

Pettit argues that human beings (as opposed to *minimally minded agents*) can comply with basic constraints of rationality by forming meta-representations [Pettit, P. 2006a. 1090]. He notes three important points related to this general capacity. They explain, in a functional sense, why we sometimes form meta-presentations to comply with basic rules of rationality. First, Pettit notes that by forming a *meta-representation* we can put "an extra-check in place of our belief-forming processes" [Pettit, P. 2007. p. 234 and 2006a. p. 1090]. So, if we believe that *p* and that if *p*, then *q*, *we* can be rationally required to form or drop a certain attitude at two different, albeit complementary, levels. At one level, we can be rationally required to believe that *q* because of our first-order beliefs (our beliefs that *p* and that if *p*, then *q*). At the other level, we can be rationally required to form (or drop) a certain attitude because of a second-order belief focused on a relation of entailment between propositions (that 'p' and 'if p, then q', taken as given, entails 'q')[150]. This second level is very important because on some occasions being required to form an attitude in the former sense is not sufficient. For instance, we can imagine cases where we are not going to believe that *q* even if we believed that *p* and that if *p*, then *q*. Surely, we could explain this phenomenon by noting that our beliefs are defined along a wide array of dispositions. We could note, for instance, the fact that certain emotional factors may be blocking the inferential dispositions constitutively connected to our

---

[150] Pettit thinks that this route works along two different paths, depending on whether we check or not our previous beliefs to form an attitude. If we do not check them, Pettit assumes that we can believe that q because we endorse a second-order belief about a certain relation of entailment. The spoken (or unspoken) expression of this deliberative route is our common *so* – meaning something close to "it is the case that q, because it is the case that p and it is the case that if p, then q'. At other times, however, we form the belief that *q* only after we have checked our previous beliefs. In such cases, the meaning of *so* is more similar to "I ought rationally to believe that q". Pettit, P. 2007. p. 234-235. See also Smith, M. 2004b for this general point.

beliefs. In such occasions, our ability to form normative beliefs could help to sustain our belief-forming processes.

Second, by forming a meta-representation Pettit notes that we can "ask questions about propositions like 'p', 'if p, then q', and 'q' – questions to do with their truth, consistency, and other relations – and to set out intentionally to let beliefs from within us in answer to such questions" [Pettit, P. 2007. p. 234. 2006. p. 143. See also Pettit, P. 2006a. p. 1095]. By doing so, we could secure the rationality of our belief-forming processes beyond the mere label of internal *coherence* [Pettit, P. 2007. p. 236][151]. We could obtain, by asking some questions about our current beliefs when we are under a MP requirement (that *p* and that if *p*, then *q*), a further level of justification for such beliefs posited in terms of consistency (with other beliefs) or truth.

Finally, Pettit stresses a common intuition to sustain his favoured account of reasoning. It is because reasoning "(…) makes its appearance when a red flags goes up, indicating that special care is needed (…)" [Pettit, P. 2007. p. 235] that we should remain faithful to the view that conceptualize it in terms of meta-representations or second-order normative beliefs. Although we can reason in a computational sort of way [Broome, J. 2006. p. 195], forming beliefs as result of other beliefs, *proper* instances of reasoning (identified by the use of expressions like *so* and *therefore* and by being embedded in contexts where 'special care is needed') are better conceptualized as involving second-order normative beliefs [Pettit, P. 2006. p. 1100].

Until here I have stressed the function of second-order normative beliefs in our reasoning. But surely, even if we have persuasively sketched the way in which forming meta-representational states could be useful to support our processes of attitude-revision, a further task remains to be addressed: How could we explain the normativity of rationality by assuming the prevalence of such processes? How could we explain the force of our requirements of rationality by assuming the centrality of second-order normative beliefs?

Pettit has a story about that too. He starts by assuming the same intuition I presented at the beginning of this chapter. In terms of it, we can explain the force of any requirement of

---

[151] Broome partially recognizes this possibility. He writes: "No doubt the direction of your reasoning may be influenced by your second-order beliefs about what you ought to believe or about what rationality requires you to believe (…) I am not suggesting that second-order beliefs directly alter your first-order beliefs, I am saying they may influence the direction of your first order reasoning (….) If you do some second-order reasoning, it doesn't replace first-order reasoning; it directs it. It has to be first-order reasoning that ultimately alters your first-order beliefs" Broome, J. (forthcoming_1)

rationality by citing the reasons that favour the state that the requirement demands. So, we can explain the normativity of rationality by paying attention to our capacity to be responsive to reasons, i.e. to our *reasons-responsiveness*. Keeping this intuition in mind, Pettit distinguishes between the reason-responsiveness of *minimally minded agents* and that of proper *agents in a full-blooded sense*. Assuming MP again, he writes:

> "(…) in the *modus ponens* example, the relevant reasons will be that p and that if p, then q. The minimally minded agent takes account of reasons, without recognizing them as reasons, when it is prompted to by its beliefs in the relevant proposition to form the belief that q. We human beings take account of reasons, treating them as reasons, when we are moved not just by those beliefs, but also by the meta-propositional belief that 'p' and that 'if p, then q' entail 'q' (…)" Pettit, P. 2007. p. 236

Therefore, the *full-blooded agent* in Pettit's sense can comply with MP because, besides of being directly responsive to the reasons offered by the fact that p and that if p, then q, he can also respond to an *attitude-relative reason,* tracked by his second-order normative belief[152]. The fact that   *his* believing that p and his believing that if p, then q encapsulate, can offer him a *structural reason* in favour of a certain pattern of attitudes. *Second-order normative beliefs* would track, or so Pettit argues, these *structural reasons*, i.e. those connected with the paramount value of *consistency*[153]. Thus, a possible chain of reasoning to comply with MP would move, accordingly to Pettit, along the following path. Once MP is applied to me because I believe that p and I believe that if p then q I will form the same *second-order normative belief* that *the received view* requires to work:

MP ᴮᴱᴸᴵᴱꜰ-ᴡꜱ B ᵢ (Oᵢ (believe that q if I believe that p and I believe that if p, then q)

---

[152] As I understand Pettit's remarks at this stage, he defends that the agent can move from a purely *deontic* normative belief to a *reason-oriented* normative belief, which he calls a *reason-providing belief.* Ordinary reasoning, Pettit argues, involves this type of beliefs but not necessary *reasons-ascribing beliefs*, i.e. beliefs giving us reasons to form an attitude as a result of our entering the scope of a basic rule of rationality.
[153] Pettit defines *structural reasons* in the following terms: "(…) structural reasons are facts about the attitudes of the agent that will require certain other attitudes to obtain – or certain attitudinal pattern to obtain" Pettit, P. 2007. p. 237. See also Scanlon, T. 2007a. p. 84-85.

After that, I will form a *reason-providing belief* departing from my second-order normative belief (my *meta-propositional belief* focused on the relation of entailment). *Reasons providing beliefs*, in Pettit's sense, are beliefs focused on the *structural reasons* that our own attitudes provide:

> **MP** **BELIEF-WS** B $_I$ (Reason $_I$ (believe that q if I believe that p and I believe that if p, then q)

If, after checking my attitudes, I find that it is true that p and that if p then q, I can move forward, forming a reason-ascribing belief:

> **MP** **REASON-ASCRIBING** B $_I$ (Reason $_I$ to believe that q because I believe that p and I believe that if p, then q)

And from a reason-ascribing belief conditionally focused on p we can form, or so Pettit assumes, a belief that p [Pettit, P. 2006. p. 1099-1100. Although see 2007a at page 501 for an important proviso about our responsiveness to second-order normative beliefs].

In sum, if Pettit were right we could offer a *general explanation* of the role played by second-order normative beliefs. Essentially, such beliefs would raise the chance of satisfying basic rules of rationality [Pettit, P. 2007a. p. 499]. As Pettit sometimes notes, they would supply an additional check for those attitudes on which our rules of rationality are applied. In doing so they would secure our reliability to comply with such rules of rationality. It is because of that - along with the minimal set of commitments such beliefs carry along - that second-order normative beliefs are plausible.

Third Task: Locating second-order normative beliefs into Kolodny's favoured requirements

To offer a general explanation of the plausibility of second-order normative beliefs is *not* equivalent to debunk Kolodny's main argument against the proposed reduction of the normativity of rationality to the normativity of reasons. As I noted before, even if Kolodny's argument is based on *the reasoning test* and such test is apparently grounded on a narrow understanding of our processes of reasoning, by simply arguing for a second-order model of

reasoning we are not debunking Kolodny's case. And this is so because of the simple fact that, while Kolodny focuses on rules of rationality framed *normatively,* Pettit constructs his case for second-order normative beliefs around rules of rationality widely understood, i.e. rules governing attitudes without an explicit normative content [Pettit, P. 2007. p. 237 and 2007a. page 500 note 4 for an explicit endorsement of this strategy]. Thus, besides of debunking Kolodny's appeal to *the reasoning test* by means of a second-order model of reasoning, we should explain further how second-order normative beliefs enable our compliance with rules of rationality narrowly formulated. Only by doing so we could reject Kolodny's case against WS theories in a conclusive way. But how feasible is this task?

In my opinion, this is a feasible task. If I am right, it simply requires that we apply the model of reasoning sketched by Pettit to those requirements framed in normative terms. After all, if normative beliefs are a special sub-set of the wider set of the beliefs we endorse, why should we not apply Pettit's insights to those requirements grounded on beliefs whose content is partially normative? Otherwise: Even if Pettit's explanation is not primary concerned with the normativity of rationality (Pettit himself confesses to be interested on the *structure* of our reasoning [Pettit, P. 2007a. at page 500]), why should we not try to use Pettit's general model of reasoning to explain also the nature of mental activity in those cases in which we fall under a rule of rationality narrowly construed?

Let us suppose that, because of your beliefs, you are exemplifying a Kolodny's type-requirement. In such situation you believe that there are reasons supporting the formation of a certain attitude (F). A basic rule of rationality could be applied to you because of such belief. If rational, you should comply with such rule by reasoning from the *content* of the psychological state on which the norm of rationality supervenes. It is because of this basic intuition that Kolodny assumes that basic rules of rationality are NS. And it is because of the fact that some rules of rationality are NS that he defends that the normativity of rationality cannot be explained in terms of reasons (remember that NR plus NS equals to *bootstrapping*). Against this by now well-known story, what I am trying to defend here is quite simple. If second-order normative beliefs are functionally and philosophically plausible – they could raise your chances to comply with any rule of rationality and we can conceive them without incurring on excessive metaphysical burdens – why do not we have yet a principled basis to fulfil Kolodny's *reasoning test* beyond of a *content-oriented* model of reasoning?

Let us suppose that the following requirement is applied to you because of your beliefs [Broome, J. 2005]:

       Krasia O (to intend F, if you believe that you have reason to F)

If Pettit were right, you could comply with this requirement by believing something like:

       Krasia $_{\text{BELIEF-WS}}$ B $_I$ (Reason $_I$ (to intend F, if I believe that I have reasons to F)

After forming your second-order normative belief, you could have an additional stance to check the reliability of your response – by checking, for instance, those first-order beliefs of yours that are supporting your normative belief. Depending on the result of that additional checking, you could either form - going forward on the basis of your second-order oriented belief and your first-order oriented checking - an intention to F or you could drop your first-order normative belief – on the basis, again, of a second-order checking being channelled by means of your first-order beliefs about the reasons supporting F. Although in this scenario your second-order belief is merely directing your first-order reasoning (and here Kolodny is right), the thing to keep in mind is that such belief could enable a better response from you to a certain kind of value, a value constitutively derived from your current mental states (*coherence*). A second-order belief, to put in Pettit's favoured words, would secure a better response to such value, a response more reliable than the one you could exemplify if you were not equipped with a capacity to form such beliefs. But then why should we remain sceptic about the powers of a *second-order account of reasoning* to restore the normative link between reasons and rationality?

Kolodny offers three main arguments against *any* account of rationality exploiting a route of reasoning other than a *content*-oriented one. These arguments, I suspect, can be formulated also against those proposals avoiding the mistakes sketched above (see footnote 144).

First, Kolodny suggests that even assuming the initial plausibility of second-order normative beliefs, in some cases you could not avoid a charge of irrationality by supplementing your reasoning with a second-order level of normative awareness, i.e. by adding a second-order normative belief. If *Krasia* can be applied on you at t_1 and you do not intend to F at t_2, then (even if you end by revising your normative belief at t_3 as result of a

second-order normative belief) you are being clearly irrational by not intending F at t_2 [Kolodny, N. 2007. p. 9 and 12].

Second, Kolodny assumes that the model of reasoning I sketched above contradicts some deep intuitions about the contexts where we are evaluate A in terms of how well she is obeying a basic rule of rationality. To Kolodny's eyes, the normal target of these evaluations is the set of A's psychological states over which the basic requirement is applied. It is because we consider a set of psychological processes around these states that we can evaluate A's rationality in obeying a requirement. In the case of *belief-closure*, for instance, our target is the set of processes composed around the *content* of the belief that p and the belief that if p, then q. In the case of *Krasia*, our targets are the processes around the content of A's normative belief about F and the intention to F. But once our common focus is fixed, it should be obvious how a second-order model of reasoning would fall apart from it. Roughly, it is because we presuppose an external state to assess A's rationality (a second-order belief focused on A's normative relations between her psychological states) that we should better reject a generalized second-order account of reasoning. And it is so because it would be very counterintuitive to evaluate A's rationality in obeying a given state by considering other states than those on which the requirement ruled its demands [Kolodny, N. 2007. p. 10].

Finally, even if these problems were solved, Kolodny notes that it is not very clear that second-order normative beliefs are always available to sustain our reasoning. At least in those cases where you cannot reason from other content than your first-order beliefs, there is only one route to comply with what is being demanded of you. In the case at hand, at the moment you ceased to have a second-order normative belief - *Krasia* BELIEF-WS – there would only be one route available for you to satisfy *Krasia.* Namely: to intend F. But having only one route is equivalent to leave the door for *bootstrapping* open again [Kolodny, N. 2007. p. 10-11].

Now, let me say something about these criticisms.

Leaving aside particular cases [see Broome, J. 2007. p. 368], it is prima facie reasonable to presuppose that the content of our normative beliefs exerts some sort of normative pressure on us by default– at least when they are understood as verdicts, as statements directly relevant to form an attitude. Kolodny assumes this pressure wholeheartedly in his first criticism. It is because of that conviction, I suppose, that he argues that those agents who ignore such pressure in a given situation (not intending F in light of their beliefs that there are reasons to F) are behaving irrationally. But why is this so? Why should we show any kind of unqualified

deference to our self-referred, *content*-oriented, normative beliefs in directing the stream of our deliberation?

Surely the answer Kolodny could give us here is twofold. Either we should deliberatively defer to our normative beliefs because *there is no alternative content or stance from which we could ground a WS process of reasoning* (I assume that this is Kolodny's core argument in *Why Be Rational?*(2005)) or we should defer to normative beliefs *because attending to them would increase our chances to act according to reasons* (even if on some occasions we do not have an objective reason in support of the required attitude). Kolodny's twofold-style answer is problematic, nevertheless. And this is so because Kolodny has envisaged a *content* from which we can sustain a wide-scope process of deliberation, i.e. he has recognized the initial plausibility of a second-order stance to ground our deliberation. If so, by Kolodny's own lights the first route available to explain our default deference to the content of our normative verdicts should be ignored. We should shift our attention then toward the second route. And these, of course, are bad news for Kolodny. The second route, remember, is about the *advantages* or *utility* of deferring to normative beliefs. And here Kolodny has no a-priori advantage over Pettit. Moreover, did not Pettit motivate his own case in favour of a second-order model of reasoning by assuming that second-order normative beliefs would increase our chances to respond to reasons?

Now, if we agree on the fact that

-   (i) The best route to exclude reasoning from second-order normative beliefs is to provide an explanation of why your chances to respond to reasons are higher by always deferring to normative beliefs

and

-   (ii) Pettit has offered an explanation at this precise level, i.e. focused on why being able to form second-order normative beliefs would increase our capacity to respond to reasons.

Why should we presuppose that a second-order model of reasoning couldn't be true? After all, Kolodny has not offered yet an argument to sustain the unqualified utility of a first-order

model of reasoning. So, what is the damage of appealing to second-order normative beliefs if we have even depicted a plausible explanation focused on their utility? And moreover, where can we find Kolodny's story to secure our capacity to respond to reasons by unconditionally following the stream of our first-order normative reasoning? Remember that Pettit depicts a story in terms of the role played by second-order normative beliefs for revising our attitudes. It would be plausible to assume, he argues, that the systems equipped with such capacity would be able to respond better to reasons. But again, where should we look in Kolodny's programme to secure his content-oriented account of reasoning?

Now let me say something about Kolodny's second point. Here I will simply suggest that Pettit's proposal, if properly unfolded, could be consistent with Kolodny's intuition about the nature and scope of our evaluative assessments of rationality. To claim this I will follow the path opened in the previous paragraph. In essence, if we are disposed to defend that we can evaluate A's responses to a requirement of rationality (R) only by assessing the quality of a certain variety of A's psychological processes (the ones structured around the *content* of those psychological states over which R supervenes), on the basis of which criterion we are excluding A's second-order normative beliefs from among the set of psychological processes that would help us to evaluate A's rationality in obeying R? As in the previous criticism, I believe that Kolodny's rebuttal at this stage would require further and detailed discussion. To say the less, it would require dealing with some questions about the type of knowledge or access presupposed by the formation of a propositional attitude. If it were proved, for instance, that a certain level of self-knowledge is constitutive to any belief or intention [Shoemaker, S. 1996- Burge, T. 1998], why should we try to narrow our evaluations of rationality to first-order oriented processes of reasoning? Do not misunderstand me. The point that I want to stress here is that, even if Kolodny disagreed with the tenability or stringency of such level of self-awareness, we should settle these issues in advance to narrow the discussion about the focus of our evaluations of rationality. As I noted before, Pettit offered a sort of tale to locate certain second-order processes of attitude-revision into a wider justificatory framework. It was an attempt to motivate the plausibility of the level of awareness I envisaged above. The question is, again, if we can find an account at this level in Kolodny's writings.

Finally, I would like to make a brief comment about Kolodny's last criticism. Here my disagreement with him is focused again on a wider topic. Roughly, I believe that the relevance

of a given scenario to understand the workings of a given concept or ability - in the case at hand, the relevance agents lacking the ability to form second-order normative beliefs to certain issues around the scope of our *ought* of rationality - is always a matter of degree, which has to be supplemented by a wider story that supports the conclusion we want to establish (see Chapter 4, section 2.1). Thus, even if we could easily imagine a situation where K can be applied and a second-order normative belief is not formed, we should not move forward to claim that second-order normative beliefs are not an essential component of our ability to respond to reasons. Rather, we should ask: Could a positive evaluation of the rationality of an agent obeying K without forming a second-order normative belief be consistent with our better account of rationality and deliberation?

4. Do we have an argument from *Meta-rationality* now?

Let us suppose that by stressing the role of *second-order normative beliefs* we could secure a WS interpretation of the *scope* of our basic rules of rationality[154]. Let us assume, in essence,

---

[154] Until now I have acknowledged the role played by first and second-order normative beliefs in sustaining processes of reasoning. However, we should also note the central role played by certain *non-deliberative dispositions* to sustain and direct reasoning (see Bratman, 1987). *Non-deliberative dispositions* to reconsider (or not-reconsider) are not over and above our unitary and partially transparent processes of reasoning, or components of certain process of reasoning but not others, or abilities involved at some particular stage of reasoning but not others. In some sense, non-deliberative dispositions are ubiquitous components of every single process of reasoning, at any stage and at any level. They complement, so to speak, first-order beliefs and second-order normative beliefs. To illustrate the role of these non-deliberative dispositions let me introduce a simple example. Let us suppose that you believe that there are reasons for a certain attitude. We can agree that such normative belief is usually based on some complex processes of evaluation of evidence, carried out via first-order beliefs and sub-agential capacities that involve certain epistemic dispositions and habits. These processes are stocked, so to speak, by the normative belief just referred. Once the belief is formed the agent will move forward to form the attitude. Nevertheless, before he can form the belief the context changes. The change is not necessary connected with the evidence or reasons grounding his previous epistemic judgment. In the new context the agent has more time in between the formation of the normative belief and the formation of his intention. Now, let us assume that the agent does not form the belief as direct consequence of the availability of more time. This simple phenomenon enacts a certain *non-deliberative disposition to reconsider*. Taking into account such non-deliberative disposition, the agent reconsiders the evidence for the normative belief. In terms of the WS framework depicted before, being non-deliberatively disposed to reconsider the evidence supporting the normative belief is equivalent to effectively dropping it. We should accept, in consequence, that the agent finishes the temporally extended process by means of which he can comply with a basic rule of rationality (the one initiated by believing that there are reasons for p) by dropping the normative belief – against Kolodny's basic insight. The agent has moved from not believing that p to not believing that there is evidence for p as result of a certain non-deliberative disposition to reconsider. Is the agent being rational here? To some people he is being rational here by dropping a normative belief because of the fact that he is expressing a non-deliberative disposition that could be rational to have in the long run. In these situations it would be possible to re-evaluate

that the normativity of rationality is reducible to the normativity of reasons only insofar we assume the role of second-order normative beliefs (sections 1.1. and 1.2). Although there is a problem with the central intuition of this chapter (section 2.2 and 2.4), as far as we are able to depict a certain process of reasoning that enable us to comply with a WS requirement of rationality (second-order reasoning), the intuition that unfolds the normativity of rationality departing from the normativity of reasons can remain in place (section 3).

Once the tenability of second-order normative beliefs to sustain our basic intuition is assumed, however, the question I want to answer goes as follows: To what extent can we obtain some insights about the psychological scaffolding that enables our mastery of basic normative concepts such as *ought* from the previous discussion around the *scope* of our basic rules of rationality? And more precisely: How could we connect such findings - located at the level of *Meta-rationality* - with the psychological nature of our moral requirements, i.e. with *meta-ethical* positions that focus on the psychological status of our moral opinions?

In what follows I will argue that even if the treatment of certain questions about the *scope* of basic rules of rationality would suggest a cognitive account at the level of *Meta-rationality*, we do not have yet a knockdown route to argue for cognitivism at the level of *Meta-ethics*. And it is so because some recent meta-ethical proposals have loosened the sufficiency of a core assumption, which is widely used to mark the boundaries of the debate between certain meta-ethical views. In terms of this assumption, it is the type of psychological state *expressed* by a moral (or normative) utterance what would give us a fundamental insight to privilege a general account of the meaning and status of our moral or normative terms [Rosen, G. 1998. p. 387-388]. Thus, if it were proved that moral sentences express beliefs, we could defend some theses about the truth of our moral judgments and their ontological status. By contrary, if it were proved that moral sentences express desires or other type of pro-attitudes we should doubt about the tenability of certain basic categories such as moral truth, moral objectivity, and moral agreement.

But as I have just suggested, supporting the taxonomical sufficiency of the category of *expression* is not free of problems. As soon as a non-cognitivist *about rationality* were able to

---

the evidence grounding a normative belief about the reasons (against Kolodny) in virtue of certain non-deliberative dispositions to reconsider. Although our process of revision is not supported from a deliberative perspective as in Pettit's account, the reverse process of reasoning is taking place in an overall rational way. In Bratman's account, a far as the agent revises his normative belief trough a rational disposition to reconsider his revision, he should be assessed as rational.

depict a sound story explaining the cognitive import of our moral and normative terms, the directness of certain insights about the psychological nature of our capacity to respond to reasons of rationality would be severely impaired. In the terms favoured by some recent non-cognitivists proposals, we could accept that *ought* – when included in sentences uttered in contexts where we are trying to *direct* our attitudes in terms of rationality - expresses a variety of second-order normative beliefs, without assuming the truth of cognitivism about *Meta-rationality*. To put it simple: by assuming that in responding to certain reasons (reasons of rationality) we are expressing a certain variety of beliefs (second-order normative beliefs) we have not proved yet the truth of cognitivism about *Meta-rationality*. And we have not arrived yet to a knockdown argument to solve the impasse in *Meta-ethics* departing from *Meta-rationality*.

4.1. The Core Assumption

In previous chapters I have been concerned with an *impasse* between two meta-ethical views, *cognitivism* and *non-cognitivism* (see especially chapter 2). I have presupposed a standard way to delimitate the contours of these meta-ethical views by suggesting that they accept opposite thesis about the psychological nature of our moral opinions. In terms of this assumption, while meta-ethical *cognitivists* usually accept that:

  C $_{psych}$ Ethical sentences express moral beliefs

  non-cognitivists defend that:

  NC $_{psych}$ Ethical sentences do not express moral beliefs

 Although both claims can be in principle formulated without paying attention to a semantic level of analysis, it is customary assumed that **C $_{psych}$** and **NC $_{psych}$** would support a semantic claim [Jackson, F. 2000. p. 10. Smith, M. 2000. p. 15-16]. In terms of this orthodox intuition, while cognitivists assume that:

  C $_{semantic}$ Ethical sentences are true

non-cognitivists should be characterized as defending that:

NC _{semantic} Ethical sentences lack truth-values

When this additional level of analysis is noted, a supplemented definition of cognitivism and non-cognitivism can be unfolded [Ridge, M. 2006. p. 2007. p. 53]. In these terms, Ridge assumes that cognitivism would accept:

C _{psych + semantic} For any ethical sentence M, M is conventionally used to express a belief (a moral belief) such that M is true if and only if the (moral) belief expressed by M is true

while non-cognitivism would be committed with:

NC _{psych + semantic} For any ethical sentence M, M is *not* conventionally used to express a belief (a moral belief) such that M is true if and only if the (moral) belief expressed by M is true

But as it is clear by these definitions, non-cognitivism needs to be supplemented further to give voice to the specific (non-cognitive) theory that is usually opposed to meta-ethical cognitivism [Rosen, G. 1998. p. 388-389. Smith, M. 2000. p. 17]. It can be done by supplementing NC $_{psych}$ with a positive claim about the type of psychological state expressed by a moral utterance:

Expressivism (NC $_{psych + semantic}$) plus the thesis that ethical sentences are conventionally used to express pro-attitudes

Thus, the polemic between cognitivists and non-cognitivists must be preliminary recasted as one established between *cognitivists* and *expressivists*. That is, a confrontation among those who defend that moral judgments express cognitive attitudes and those who assume that moral judgments are conventional devices aimed to give voice to non-cognitive, action-guiding, attitudes (*emotions*, *universalizable prescriptions* or *states of norm-acceptance*).

But if the debate is framed in this way, a certain possibility is tacitly ignored by both meta-ethical views. As Michael Ridge has recently reminded us [Ridge, M. 2007. p. 52-53], by positing the debate along the previous lines we hide the following possibility:

> **Ecumenical Accounts** Ethical sentences are conventionally used to express both (moral) beliefs and desires or pro-attitudes.

Although it is difficult to explain why ecumenical accounts have been so widely neglected in recent discussions, it must be accepted that their concealment is one of *the core assumptions* underlying the current debate between cognitivists and non-cognitivists. It goes without saying that when we defend a meta-ethical position about the psychological import of our moral opinions, we are referring either to a variety of *cognitivism* or to a variety of *expressivism*, excluding from the very beginning the possibility of an ecumenical account. But how could we explain such widely assumed concealment? And moreover: How is it related with the strategy I have favoured all along this chapter?

Let me start by offering a brief answer to the first question. Although I have sketched before a general framework that would help us to accomodate an eccumenical approach (see Chapter 2, section 3) the answer I will support here is basically the one proposed by Terry Horgan and Mark Timmons. Surely, this option would be a good standpoint to connect issues regarding the status of morality with other issues related to the status of rationaliy [see especially Horgan, T. Timmons, M. 2000/2007 and 2006a. Although see Skorupski, J. 2001].

These philosophers start by distinguishing three different levels of semantic content. Let us suppose, they assume, that A utters the following sentence:

(1) Stealing money is wrong

The grammatical form of (1) is declarative so (1) is a declarative sentence and the judgment expressed by (1) has ***declarative content***. As typical moral judgments mirror the grammatical form of (1), *any* meta-ethical theory must assume that moral judgments have declarative content simply "as a result of grammatical form" [Horgan, T. Timmons, M. 2000/2007. p. 58].

But even if (1) has *declarative content* as a result of its grammatical form, Horgan and Timmons note that there is a further disagreement about whether such content implies by itself that moral judgments have also ***cognitive content***. Or to put it another way: they note that there is a deep disagreement about whether the level of semantic analysis facilitated by

certain *grammatical* features involved in the utterance of a moral sentence can directly fix the *psychological* content we express in uttering such sentence.

Clearly this is the precise level of disagreement where we located before the dispute between cognitivists and expressivists. About this precise level, Horgan and Timmons endorse an essential claim:
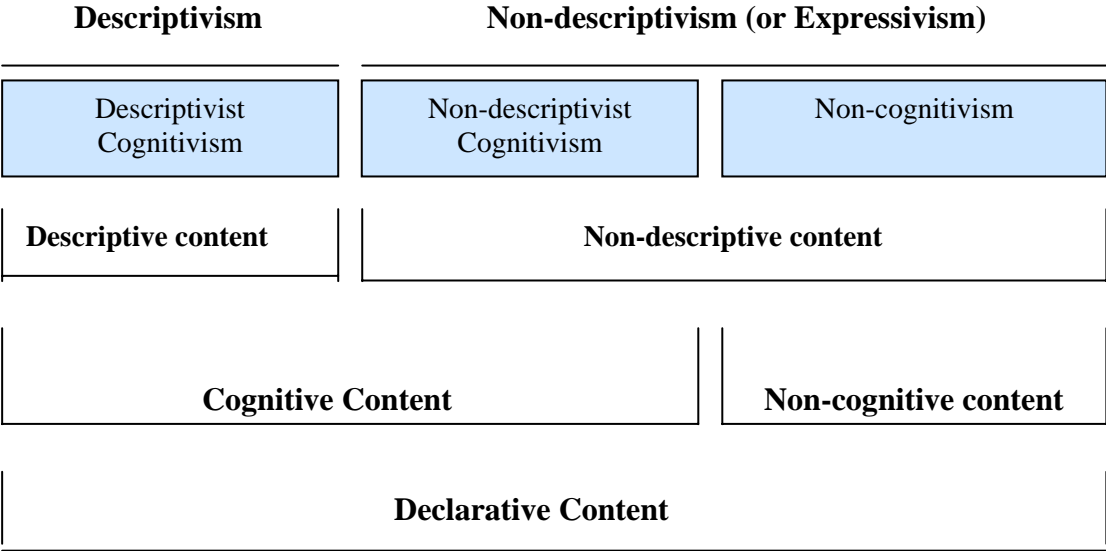
> " (…) what has been taken for granted in analytic philosophy generally, and metaethics in particular, is the idea that for content to be genuinely cognitive it must be in the business of purporting to represent how the world is" Horgan, T. Timmons, M. 2000/2007. p. 59

Thus, a third level of analysis of (1) aimed to determine its overall semantic content should be focused on whether (1) has **descriptive content**. And it is at this precise level that the core assumption in *Meta-ethics* emerges. It assumes that if (1) has cognitive content then (1) must be necessary analyzed as having *descriptive* content because any cognitive content purports to describe some realm of facts [Horgan, T. Timmons, M. 2000/2007. p. 59 and 2006a. p. 230]. It is because (i) *cognitivism* at the level of psychological content is customary conflated with *descriptivism* and (ii) the fact the some people assumes that moral sentences do not describe any realm of facts, that we tend to accept that any *expressivistic* theory must be necessary understood as a version of *non-cognitivism*. But the fact is that if we accepted that not every belief is aimed to describe an independent realm of facts we could still conceive a *cognitive* version of *expressivism*, one in which the psychological state expressed by the utterance of a moral sentence can be partially equated with a non-descriptive belief whose basic function could be primary analyzed in terms of reasoned action-guidance [Horgan, T. Timmons, M. 2006. p. 233][155].

---

[155] In order to explain why irrealist cognitivism is widely ignored, Skorupski writes: "Such a view seems at first sight attractive, and sensible, yet it is surprisingly uncommon. While there may be various reasons for that, I think there is one very fundamental reason. A deep metaphysical obstacle seems to lie in its way: informative cognitive content just seems to *be* factual content; content is factual content. The cognitive irrealist will have to dispel the seeming force of this idea. For if all cognitive content is factual content in the intended more than merely nominal sense then, it seems, we must be either non-cognitivists or realists about normative claims". Skorupski, J. 2000. p. 117-118

Clearly, once the unquestioned appeal of *descriptivism* at the level of psychological content is removed, the standard meta-ethical taxonomy must be put under a different light. To start with, when the influence of *descriptivism* at the level of content is fully unfolded, we cannot assume a direct partition between *cognitivism* and *non-cognitivism*. Instead, we must accept that the division is rather between semantic *descriptivism* and semantic *non-descriptivism* or *expressivism*. In terms of this additional taxonomy, it can endorsed an *expressivist* approach to the nature of our moral judgments without renouncing to *cognitivism* – and without rejecting that we assert something when we utter a moral sentence. To do so, we should simply reject the intuition saying that all beliefs describe how things are, offering, additionally, a framework to understand the role played by non-descriptive beliefs in our deliberation [Horgan, T. Timmons, M. 2000/2007. p. 61-62. 2006a. p. 232-233].

To put it graphically [adapted from Horgan, T. Timmons, M. 2000/2007. p. 60]:

Descriptivism	**Non-descriptivism (or Expressivism)**	
Descriptivist Cognitivism	Non-descriptivist Cognitivism	Non-cognitivism
Descriptive content	**Non-descriptive content**	
Cognitive Content		**Non-cognitive content**
Declarative Content		

An important consequence to be extracted from this re-formulation of the debate is that once the current polemic in *Meta-ethics* is modelled along with Horgan and Timmons's favoured schema, we have conceptual space to systematically investigate the possibilities of ***ecumenical accounts*** in *Meta-ethics*. As I noted above, any ecumenical account would accept the following claim:

Ecumenical Accounts Ethical sentences are conventionally used to express both beliefs and desires.

In a recent defence of *Ecumenical Expressivism*, Michael Ridge offers a succinct formulation of the two possible varieties to be found in *any* ecumenical account of the psychological nature of our moral opinions. In Ridge's terms [compare with Ridge, M. 2007. p. 54]:

Ecumenical Cognitivism: $C_{semantic + psych}$ plus the thesis that ethical sentences are conventionally used to express pro-attitudes.

Ecumenical Expressivism: $NC_{psych + semantic}$ plus the thesis that ethical sentences are conventionally used to express beliefs.

The difference between these versions of the ecumenical genus is made clear by Ridge himself when he writes:

> "The Ecumenical Cognitivists and the Ecumenical Expressivist agree that normative utterances expresses both beliefs and desires. They disagree about the connection between the truth of the belief expressed and the truth of the sentence which expresses it. The cognitivists insist that a given normative sentence is true if and only if the belief that it expresses is true, whereas an expressivist denies this" Ridge, M. 2007. p. 54

Leaving aside the sense in which ecumenical theories still disagree about the status of our moral opinions even if they accept a basic parity at a psychological level, the important point is that once we assume that there is conceptual space for ecumenical versions of expressivism and cognitivism, the overall project defended in this chapter must be reassessed. Because, how should we interpret the findings coming from a second-order analysis of the way in which we respond to reasons of rationality if we assume the conceptual tenability of these ecumenical accounts? If the general point stressed by Horgan and Timmons were feasible,

merely noting that we can respond to certain reasons of rationality through the exercise of a capacity to form second-order normative beliefs does not establish the truth of cognitivism at the level of *Meta-rationality*. As the previous diagram suggests, when the meta-ethical debate is recast in the way favoured by Horgan and Timmons, we could assume that the ought of rationality is cognitively loaded without reaching an agreement about the proper semantic category exemplified when we appeal to the ought of rationality in contexts of reasoning and deliberation. If we wanted to establish a more substantive conclusion, maybe we should investigate how the truth-conditions of ought in these first-personal contexts of deliberation are fixed. Additionally, we should determine the function and deep rationale that underlies the semantic notion of *expression*, and whether it is constitutively connected to certain cognitive processes or (despite of depending on the well-functioning of certain cognitive processes) it can be understood in non-descriptive terms. These additional tasks would help to settle down the debate about the type of psychological state underlying our capacity to respond to reasons of rationality. They would clarify, so to speak, the sense in which asking second-order questions *about rationality* would be helpful to solve core meta-ethical disputes. In the next section I will highlight these questions, assessing the relevance of the revised taxonomical partition of the meta-ethical domain for what I have labelled in the current chapter as *Meta-rationality*.

4.2 Applying the Core Assumption at the level of *Meta-rationality*

If the main point defended in the previous section is plausible we would not have a sufficient basis to cancel the *impasse* surrounding *Meta-ethics* by asking questions about *Meta-rationality* unless we are able to:

(1) Specify the main differences between ecumenical versions of *cognitivism* and *expressivism* at the level of *Meta-rationality*.

(2) Explain to what extent the previous findings about the psychological scaffolding that enables us to comply with basic rules of rationality are better explained by cognitivism rather than by an ecumenical version of expressivism about *Meta-rationality*.

A good reference point to note the importance of (1) and (2) at this stage is Allan Gibbard's recent meta-ethical proposal. In this way we could understand better how even if meta-propositional beliefs could explain the way we respond to reasons of rationality, the debate about the psychological status of our basic normative concepts would not be fully closed in favour of a certain version of cognitivism about *Meta-rationality* unless we offer some answers to the above questions [Pettit, P. 2006. See Blackburn, S. Sinclair, N. 2007. Chrisman, M. forthcoming. See Gibbard, A. 2007 and Scanlon, T. 2007 for a recent exchange about Gibbard's last book]

I introduced the main lines of Gibbard's theory before (see chapter 2, section 2.2.). Remember that, in Gibbard's terms, the basic interest *Meta-ethics* has is psychological:

> "The expressivist strategy is to change the question. Don't ask directly how to define 'good' (…) Instead of a straight definition, expressivists propose, seek a characterization of a different form. Ask what *states of mind* ethical statements expresses" Gibbard, A. 2003. p. 6 (Gibbard's emphasis)

The other basic feature of Gibbard's account is his belief on the utility of a two-stage approach to highlight the basic psychological question about our moral terms. In essence, he claims that to determine the psychological status of our *ethical* terms we should ask for the type of psychology exemplified by an agent who uses basic normative terms, i.e. for the psychology exemplified by a thinker-planner using judgments of *what to do*:

> "My ultimate aim is to study the workings of familiar ought-laden or 'normative concepts, like *good*, or *admirable*, or *reprehensible*, and offer a expressivist theory of them. To do this, though, I start out considering just one particular kind of judgment, a kind for which expressivism *must* be right. These are judgments of what to" Gibbard, A. 2003. p. 7 (Gibbard's emphasis)

Once these basic points are assumed [see Pettit, P. 2006. p. 1083], Gibbard's semantic insight about our judgments of *what to do* goes as follows:

> "(…) to conclude, say, that fleeing the building is the thing to do just is to conclude what to do, to settle on fleeing the building (…) If I assert 'Fleeing is the thing to do', I thereby express a state of mind, deciding to flee"
>
> Gibbard, A. 2003. p. 7-8

Thus, the predicate *is the thing to do* is a device to express decisions [Gibbard, A. 2003. p. 41]. Decisions are sometimes the results of *planning states* or *plans* [Gibbard, A. 2003. p. 48. Pettit, P. 2006. p. 1084]. In Gibbard's sense, a *plan* will allow the agent to form a view about what to do in an *actual* situation [Gibbard, A. 2003. p. 53]. But a *plan* will also help to decide what to do in certain *contingent* scenarios and even in *hypothetical* cases "we know won't arise" [Gibbard, A. 2003. p. 52]. Moreover, a *plan* about other person's situation could also be formulated along any of the previous dimensions [Gibbard, A. 2003. p. 50][156].

But *decisions* are not isolated acts. As far as the *plans* that determine our decisions fall under certain constraints derived from their inferential relations with other plans and beliefs - as far as our plans normally exemplify certain patterns that could be described in terms of *consistency* or *coherency* [Bratman, M. 1987][157] - we can evaluate the rationality of our decisions in terms of how the plans that lead to them respect these relations of *coherence* and *consistency* [Gibbard, A. 2003. p. 45. Pettit, P. 2006. p.1090].

Taking these previous remarks for granted, the predicate we are concerned with now - *is the thing to do* - is understood by Gibbard as an expressive device aimed to identify an alternative that is *uniquely permitted* by an agent's planning states, i.e. an alternative that is required by an agent's plans once they have been corrected by basic constraints of consistency and coherency. To put it differently: the predicate *is the thing to do* is a device aimed to identify, by expressing it, a decision that is rationally required by a given set of an agent's planning states. Thus, the concept expressed by our target predicate works, or so Gibbard argues, *expressivistically*: our mastery of such concept cannot be explained by referring to the

---

[156] Gibbard writes: "(…) a plan will be a determination of what to in various contingencies, expected and hypothetical" Gibbard, A. 2003. p. 53

[157] I will ignore here Bratman's case for *stability*

belief guiding our use, which tracks a certain type of fact, but rather by the way the predicate avows a certain decision or plan [Gibbard, A. 2003. p. 5].

Gibbard contemplates three basic advantages of having an expressive predicate with the logical behaviour of *is the thing to do*. First, he notes that this could help to check our decisions by locating them in a public context. So, if by saying that ϕ-ing *is the thing to do* I express my decision to do ϕ and doing ϕ is not in fact required by the agent's planning states, then the fact that we can avow our decisions by means of a predicate would facilitate the public checking of such decisions. To some extent, we can disagree with others about our plans in a given context because we previously made them public by using a predicate such as *is the thing to do* [Gibbard, A. p. 49].

Second, Gibbard suggests that our target predicate is also useful because it would improve our first-personal processes of deliberation. Such predicate would explicitly capture, Gibbard notes, the power of logic to regulate our psychological states [Gibbard, A. 2003. p. 13]. By returning to the previous case: if I say that φ-ing is the thing to do and I also say that ϕ-ing is the thing to do (believing at the same time that if ϕ then γ), it seems natural to say that I should revise my decision to do ϕ simply because it is against a basic rule of logic (once we take into account my beliefs) [Gibbard, A. 2003. p. 41-46. Pettit, P. 2006. p. 1096]

Finally, to count on with a predicate such as *is the thing to do* will be important for purposes of psychological self-management. Surely, by predicating that ϕ-ing is *the thing to do* an agent could achieve a greater degree of control to secure his intention to ϕ, improving his performance with respect to the foreseeing factors that could make the formation of an intention to do ϕ difficult [Pettit, P. 2006. p. 1094]

So far we have explained the intrapersonal and interpersonal functions of a predicate such as *is the thing to do*. However, the point about Gibbard's target predicate I am concerned with is not merely focused on the personal and interpersonal advantages that it carries. Rather, I am interested in its semantic status. Or more precisely, I am concerned in how Gibbard's semantic treatment of such predicate affects the tenability of classic *expressivism* once we have assumed the conceptual possibilities open by ecumenical accounts. Under a certain interpretation, Gibbard's overall approach would suggest that further work is needed to separate *cognitivism* from *expressivism* at the level of the target predicate.

In order to illustrate the problem I have in mind, let us convene that core predicative uses are guided by certain beliefs. Let us assume, in essence, that we can properly predicate P of x

only if we *believe* that x exemplifies the properties P stands for. If this image were sound, Gibbard should accept that, as far as *is the thing to do* is behaving like a predicate, when we apply it to a certain act-type we are expressing a belief, which is guiding our predicative use. Gibbard, surprisingly, assumes precisely that [Gibbard, A. 2003. p. 76-77, 93, and specially at page183. See also Sinclair, N. Blackburn, S. 2007. p. 703]. But then an obvious question emerges: in what sense does Gibbard still endorse an unqualified *expressivistic* treatment of our target term, i.e. an expressivistic explanation of the meaning of *is the thing to do*? [Jackson, F. Pettit, P.1998. Smith, M. 2005 for this general point]

At this stage, the importance of the core assumption highlighted in the previous section emerges as an essential tool to understand Gibbard's *expressivism* about *is the thing to do*. If we note, with Horgan and Timmons, that *not* every belief is necessary descriptive, is there any inconsistency in assuming that we assert something by saying that ϕ-ing *is the thing to do* (expressing a certain belief with such assertion), while denying, at the same time, that we are describing some property out there by what we are saying, that is, by asserting that ϕ-ing *is the thing to do*? Let us focus on a simple decision to illustrate what is going on here. Let us suppose that as result of your plan to do ϕ and your belief that if ϕ then γ you assert that γ *is the thing to do*, expressing your decision to do γ. If the cognitivist account depicted by Pettit in the previous discussion were right [see Pettit, P. 2006. p. 1096], we should convene that by asserting such sentence you would be expressing the following belief:

> **(Cognitivism)** B $_I$ (Reason $_I$ (intend to do γ if I intend to ϕ and I believe that if ϕ, then *γ*)

This is a *plan-relative* or *plan-generating* belief [Pettit, P. 2006. p. 1102]. But Gibbard, as we have noted, also accepts that when you say that γ-ing *is the thing to do* you are expressing a *plan-laden* belief by means of your *assertoric* speech-act:

> **(Gibbard's expressivism)** B $_I$ (Plan-laden rational $_I$ (intend to do γ if I intend to ϕ and I believe that if ϕ, then *γ*)

This is *a plan-laden* belief [Gibbard, A. 2003. p. 93[158]]. The concern I noted before about Gibbard's account emerges thus in a clearer way if we exemplify it by referring to simple decisions. The thing to ask about these decisions is clear: if both, *cognitivists* and *expressivists* defend that, in these contexts, our *thing-to-do predications* express second-order beliefs that Gibbard conceives as focused on mixed *facts-plans* states, is there any difference between *cognitivism* and *expressivism* at the level of *Meta-rationality*?

Sinclair and Blackburn answer this question in the following way:

> "The difference between Gibbard's view and those that more commonly goes under the label of 'descriptivism' is in their explanation of our practice of making normative judgments. Descriptivism hold that some part of the explanation of that practice must involve seeing normative judgments as tracking or attempting to track some sort of property (…) Expressivists deny that the explanation of the practice will involve any such component" Sinclair, N. Blackburn, S. 2007. p. 703-704

And the answer above, which stresses an *explanatory* difference, seems to be also Gibbard's own solution [Gibbard, A. 2003. p. 183]. If Gibbard is right both, descriptivists and expressivists, would accept in principle that we express a certain type of belief when we say that something *is the thing to do* - or, as Pettit defends, "that a certain action is rational relative to one's own mental states" [Pettit, P. 2006. 1096]. The difference is that while in Gibbard's case the *expressive* act of speech makes intelligible the role of the belief being expressed (a non-representational belief by Gibbard's lights, one whose functional role is related with intra-personal and inter-personal coordination)[159], in *any descriptivist* account the belief literally tracks a certain property – either a purely normative property, or a natural property or, as Pettit argues, a property based on the rational support derived from our psychological states, i.e. the property X has of being a plan relatively rational in respect to Y,

---

[158] Horwich writes: "Just as 'x is white' is the standard expression of a belief that stems from a certain experience, 'x is rational' is the standard expression of a belief that stems from certain pro-attitudes. In both cases we suppose, in addition, that the terms are logical predicates, in both cases that their application is constrained by the speaker's states of mind (…)" Horwich, P. 1993. p. 145

[159] *Plan-laden beliefs*, Gibbard notes, are beliefs expressed from a decisional locus whose function is to avow a decision or plan. See Gibbard, A. 2003. p. 8.

where 'Y' stands for a set of mental states respecting deductive or inductive constraints [Pettit, P. 2006. 1100 and 1103][160].

But if this were so, by simply showing that a second-order belief is expressed in some simple cases of deliberation, we would *not* settle down the debate about the status of our normative terms at the level of *Meta-rationality*. Insofar as some *expressivists* contemplate the possibility of non-descriptive beliefs (plan-laden beliefs in Gibbard's sense) an additional level of *disagreement* is found at the very level of *Meta-rationality*. The disagreement is now about the status of our *representative states*, about the notion of *representation* involved in our appeal to second-order normative beliefs, and about the best way to explain our representative capacities. Are these *beliefs* supporting a descriptive reading of our general capacity to be rational (one to be explained by appealing to the same capacities we exemplify in forming descriptive beliefs) or, on the contrary, are these beliefs, if properly understood, supporting a more sophisticated case for expressivism, now at the level of *Meta-rationality*?

If I am right, we could say something positive about this problem by choosing between two different routes.

On the one hand, we could follow a theoretical-rooted path.. Roughly, we could argue that when descriptivists about rationality assume a variety of second-order normative beliefs to enable our capacity to respond to reasons of rationality, the sense of *belief* that they presuppose is not the same than the one accepted by expressivists [Gibbard, A. 1990. p. 117-120. Sinclair, N. 2006. p. 254. Cuneo, T. 2007. p. 56-57]. Evidently, for this strategy to work, we need a general notion of *belief* incoated into a general theory of *representation*. Thus, we could start by noting that our commonsensical notion of *belief* and *representation* is structured around a set of core features – *aboutness* or *phenomenological externality*, *inferential role*, *assertive import*, *truth-assesibility,* etc. After that, we could assume that any appeal to belief or representation should respect these features – as far as they define our common view on the nature of belief. Now, let us suppose that the particular appeal to *belief* favoured by expressivists is not in agreement with some of the most central intuitions about *belief* and *representation* [Cuneo, T. 2006a. p. 40-47]. If this were so the accomodation that expressivism reaches by appealing to such quasi-beliefs would be in danger. In essence, we could affirm that *expressivists beliefs* should not be considered as full-blown beliefs because they do not fulfill some core features associated with our common view of *belief* and

---

[160] In Pettit's terms, "thing-to-do-predications express plan-generating beliefs in the plan relative rationality of certain option". See Pettit, P. 2006. p. 1099

representation. Therefore, by offering an argument at this precise level – one closer to issues in *Philosophy of Mind* and *Theory of Representation* - we could re-evaluate the relevance of the just referred findings at the level of *Meta-rationality*. As far as commonsensical beliefs are defined by X, Y, and Z, and plan-laden beliefs cannot be described in these terms,we should agree that there is no *impasse* at the level of *Meta-rationality*. *Plan-laden beliefs* simply are not *beliefs*.

On the other hand, we could overcome the *impasse* at the level of *Meta-rationality* by focusing, again, on our processes of reasoning. In a certain way, by working further on the kind of normative regularities involved in such processes we could try to eliminate the notion of *belief* as a theoretical framework to explain our responses to basic requirements of rationality. If we could prove, for instance, that quasi-beliefs and quasi-representations are somehow excluded from some core cases of deliberation (because of the normative structures that certain demands of rationality posit) then we could attain an additional stance to support cognitivism about *rationality*. And the same could be applied, of course, to support expressivism at the level of rationality. In a certain way, this strategy resembles a recent approach to self-knowledge favoured by Tyler Burge. In this approach, the first-person pespective and the first-person concept developed in such perspective could be vindicated by appeal to the nature of our processes of reasoning [Burge, T. 1998. p. 248-251. See also Schroeter, F. 2005].

In the next section I will be only concerned with the second route. Although the first route is quite promising to decide the issue on the precise status of our second-order normative beliefs, I tend to believe that if you are working on practical issues the path to follow is the one asking for the kind of constraints posited by our reasoning (understood as an autonomous activity) on those descriptions of the capacity we exemplify in responding to our reasons of rationality. Even if the philosophers working on topics around *representation* and *belief* could offer a general argument to decide the issue between cognitivism and non-cognitivism at the level of *Meta-rationality*, I am convinced that if we want to preserve a minimal and autonomous space for *Meta-ethics* then we should delve into the nature of reasoning and the normative structure of our rules of rationality. By doing so, we could offer a new and original perspective into the wide debate on moral objectivity.

4.3. Reasoning, Belief and the Phenomenology of Neccessity

The case we are concerned with is one where A's beliefs at t_1 ground a deontic requirement on A's beliefs at t_2. A *deontic* requirement is one you cannot disscount in your deliberation because it offers you a conclusive reason in support of a certain attitude or pattern of attitudes. *Belief-closure* (BC) is a deontic requirement. To put it schematically:

BC$_{WS}$ If A_B at t_1 that p and that if p then q, then A *ought* at t_2 (to believe that q if A_B that *p* and that if *p* then *q*)

The right-hand side of this conditional expresses a *Wide-Scope* reading of BC. As I have argued, BC $_{WS}$ should be separated from a *Narrow-Scope* formulation of BC, which assumes that

BC$_{NS}$ If A_B that p and that if p then q at t_1, then A *ought* to believe that q at t_2

Some *cognitivists* think that the normative force of BC is properly unfolded by appealing to reasons because A could comply with BC$_{WS}$ by means of a second-order normative belief focused on the requirement as it applies on A's beliefs

BC $_{second\text{-}order\ belief}$ B_I (O_I (believe that q if I believe that p and I believe that if *p*, then *q)*

This *second-order normative belief* supervenes over A's first-order beliefs (that p and that if p then q) and over a second-order belief of A whose content is that she believes that p and that if p then q. As I noted, to some people the ability to form a second-order normative belief is *reliably* connected to the effective formation of any of the attitudes permitted by BC$_{WS}$. By forming a second-order belief such as BC $_{second\text{-}order\ belief}$ we simply believe that a (conditional) normative property is the case and that *we* exemplify such a property. To some cognitivists about rationality, this second-order normative belief is sufficient to secure our believing any of the states required by BC when we believe that p and that if p, then q [Shoemaker, S. 1996. p. 34. Jackson, F. 2001. p. 102].

As I noted before, sophisticated expressivists are happy to accept that a second-order normative belief, a quasi-belief in Blackburn's sense, is normally present in simple cases of deliberation

BC $_{\textbf{plan-laden}}$ belief B $_I$ (*Plan-laden rational* $_I$ (believe that q if I believe that p and that if then q)

The type of belief stressed here, however, is described in different terms by those committed with expressivism. Let us focus on Allan Gibbard again. In Gibbard's sense (at least in the sense extracted from some passages of *Thinking How to Live*), the paramount feature of the above-referred belief is that it does not mirror a normative property. If Gibbard is right, the belief expressed by BC$_{\text{plan-laden belief}}$ is mainly an useful way to channel your acceptance of a *policy* demanding of you to believe that q in presence of a certain set of beliefs (that p, that if p then q). Even if you are able to avow such commitment by forming a second-order belief, you are *not* exemplifying a full-blown normative belief. And this is so because your belief could be explained in the end by appealing to other mechanisms than the ones that normally cause your factual beliefs. These mechanisms are, of course, social mechanisms, which are responsible of coordinating us as social animals. Thus by forming a second-order normative belief in Gibbard's sense, you could achieve two basic goals. First, you could help others to coordinate with you. It is because others know your epistemic policies that they can predict your behaviour, adjusting their own performance to yours. Additionally, by voicing your commitment you could expose your mental life to public scrutiny (in Gibbard's sense, the public domain is the domain where, among many others things, others are giving you some advice about the quality of your attitudes). They help you, for instance, to narrow the conditional commitment that is the content of BC $_{\text{plan-laden belief}}$ by giving to you additional information about whether it is the case that p, or that if p, then q. As I have just said, both types of advantages would help to explain the function and rationale of BC $_{\text{plan-laden belief}}$.

Now, let us suppose that second-order normative beliefs can be construed as quasi-beliefs (Blackburn), as commitments with a certain epistemic policy (Gibbard), or as non-descriptive beliefs (Horgan and Timmons). Is there then any additional route beyond general issues on *belief* and *representation* to settle the issue between BC $_{\text{plan-laden beliefs}}$ and BC $_{\text{second-order beliefs}}$?

Otherwise: Even if we assumed that BC $_{\text{plan-laden beliefs}}$ and BC $_{\text{second-order beliefs}}$ are equally plausible to explain our responsiveness to reasons of rationality, is there any chance to resolve the *impasse* by turning our attention to the first-personal standpoint we occupy when we obey a basic rule of rationality? I believe that we have at least two obvious routes to overcome the *impasse* at the level of *Meta-rationality*.

On the one hand, we could try to reject BC $_{\text{plan-laden beliefs}}$ by claiming that BC $_{\text{plan-laden beliefs}}$ cannot support the kind of *stability* that is required to enable us to comply with a requirement such as BC through a WS-process of reasoning. It is important to note that in discusing practical reasoning we are familiar with a sense of *stability* which refers to a stable *content,* i.e. to a content to be shared and consistently referred to in paradigmatic processes of attitude-revision. To illustrate the distinction, let me return to BC. Remember that I have assumed that in obeying BC we form a second-order belief. Leaving aside by now the precise reading of such second-order normative belief (whether it is better understood as BC $_{\text{plan-laden beliefs}}$ or as BC $_{\text{second-order belie}}$), I suggested that we comply with BC by checking our first-order beliefs. Second-order beliefs are critical to enable us to carry out this checking. Certainly, all this process requires a certain degree of *temporal stability*. Only because of it we can sustain, from t_1 to t_2, a certain process of attitude-revision. But even more important is the shared or stable *content* required to guide this process, which governs our first-order checking. And as I see the point, those who question BC $_{\text{plan-laden beliefs}}$ could simply presuppose that our second-order beliefs are better understood as BC $_{\text{second-order beliefs}}$ because only such type of beliefs would be able to sustain the required stability for our reasoning.

But this is simply wrong, or so expressivism claims. To expressivists, the relevant criterion to determine the *stability* of a given *content* is fixed by the possibility of *agreement* on such content. Insofar as the content of decisions, plans, and commitments can be questioned, agreeing or disagreeing on it, expressivists are happy to assume that such attitudes are stable in the relevant sense, i.e. in the sense that can conform the locus of our second-order normative beliefs. Thus, as far as (i) *stability* is crucial to the process of attitude revision we are concerned with, (ii) this property is dependent upon the possibility of agreement and disagreement, and (iii) we can agree and disagree on beliefs but also on decisions, we should conclude that we cannot appeal to stability to rule out BC $_{\text{plan-laden beliefs}}$ as a feasible interpretation of the status of our second-order normative beliefs [Gibbard, A. 2003. p. 65-71].

On the other hand, we could turn our attention toward some phenomenological or experiential features attached to our deliberative standpoint. By doing so we could use them to decide the issue between expressivism and descriptivism at the level of *Meta-rationality*. We could argue, for instance, that if BC $_{\text{plan-laden beliefs}}$ were incorrectly placed to accommodate our phenomenology about deliberation, we would be entitled to reject BC $_{\text{plan-laden beliefs}}$. And the same would go for BC $_{\text{second-order beliefs}}$. Any meta-normative account going against our experience in first-personal contexts of deliberation should be discarded.

Although this strategy is attractive, I will assume here that it is not plausible overall. We cannot pretend to resolve the *impasse* at the level of *Meta-rationality* by deferring on this strategy[161]. At the very end, I think, we could gain an additional stance to distinguish better between descriptive and non-descriptive beliefs. We could not offer a knockdown argument in support of expressivism or descriptivism about *Meta-rationality* by shifting our attention toward phenomenological issues about deliberation, anyway.

The way I would like to illustrate this second route is by depicting a tentative, far-fetched, argument. At the beginning of this argument I start by assuming a platitudinous fact about deliberation. I accept that we all experience simple instances of deliberation under objectivist feelings. Once this simple fact is noted, it sounds plausible to endorse a double conditional: if expressivism about rationality were not able to accommodate our core feelings about deliberation then we should assume that expressivism is not explanatory on a par with descriptivism. If, on the contrary, it turns out that expressivism about rationality can accommodate our core phenomenological features about deliberation, then we should accept that expressivism is on a par with descriptivism also at the level of deliberation. From here I move forward to argue that expressivism can in fact accommodate our core feelings about deliberation. However, I suggest that the way in which expressivism is able to do so is not free of burdens. In a nutshell, expressivism would be able to explain our objectivist feelings about deliberation only by (i) endorsing a commitment with an error-theoretic account of

---

[161] Let me note in advance that the point I want to raise is not entirely new. It can be found in Wright, C. 1987. p. 47-48. Gibbard, A. 1990. p. 153-170. Blackburn, S. 1992. p. 346-348. Shafer-Landau, R. 2003. p. 27-30 and Brink, D. O. 2007. p. 269]. In addition, it should not be confused with Dorr, C. 2002 and Lenman, J. 2003, although both of them refer to it least tangentially. Even if not entirely new, nevertheless, I believe that we could gain a better understanding of certain questions about objectivity by formulating them at the level of first-personal deliberation – see Horwich, P. 2005. The general line of argument is also familiar in the debate about free-will. See Watson, G. 2004/2003. Velleman, J. D. 2000/1989. Nahmias, E. et al. 2004 as cited by Holton, R. 2006.

deliberation or by (ii) offering an account of our deliberative perspective *of explanatory typ*e other than the one sketched by descriptivism. Insofar as expressivism has consistently rejected any error-theoretic account of our normative capacities, we should accept that expressivism offers, in fact, an explanation of our responsiveness to reasons of rationality *of an explanatory type other than the one favoured by descriptivism.* Thus, even if by focusing on issues about deliberation we cannot resolve the *impasse* between expressivism and descriptivism at the level of *Meta-rationality*, it should be clear that by locating the opposition at the first-personal level of deliberation we could better grasp how the real controversy is, as Blackburn noted above, about explanation. What are we disposed to accept as a correct explanation in order to make sense of our deliberative capacity to be rational? Let us ask this question, I suggest, and we could have an additional standpoint to resolve the impasse in *Meta-ethics*.

If Gibbard is right, by believing the content expressed by BC $_\text{plan-laden beliefs}$ an agent is expressing a conditional commitment, ruling-out a certain scenario (one where he would fail to believe that q while still believing that p and that if p, then q). As I noted above, the agent envisaged by Gibbard does not believe the content of BC $_\text{plan-laden beliefs}$ strictly because of the content of his own first-order beliefs. On the contrary, he believes the content of BC $_\text{plan-laden beliefs}$ only *when* he exemplifies a certain set of beliefs [Smith, M. Stoljar, D. 2003]. He does not believe such content simply *because* of the fact that he is exemplifying a certain set of first-order beliefs – as a plan-relative believer in Pettit's sense would do. In Gibbard's plan-laden terminology, by believing the normative content posited by BC $_\text{plan-laden beliefs}$, the agent is voicing a commitment or plan (a stable policy) whose deep rationale should be understood in terms of either personal self-management or public criticism[162].

Things are different, however, if we endorse a realist, truth-functional construal of *necessity*. In such account, if it is true that p and that if p then q, then it is *necessary* true that q. Consequently, if you believe that p and that if p then q, then you necessary ought to believe that q. By appealing to a second-order normative belief, therefore, realists would assume that you are simply helping yourself to check whether this type of necessity is operative in your mental states. They could even allow us to formulate our rules of rationality to be sensitive to

---

[162] Jackson describes this position along the following lines: "The idea would be that instead of thinking of

If X believes that P, and X believes that if P then Q, then X ought to believe that Q

as saying that anyone that satisfies the description 'believes that P and believes that if P then Q', also satisfies the description 'ought to believe that Q' , we should think of it as prescribing a constraint on the attitudes that can properly be taken". Jackson, F. 2001. p. 108

this fact, as wide-scope rules. Anyway, the crucial point for practical realists is that the kind of *necessity* we are trying to accommodate in these simple cases of deliberation is somehow fully independent of whether you are committed or not with a certain policy. To deliberate, at least to deliberate in realists' favoured sense, is to find out whether you fall under this type of deontic *necessity*, an *externally* grounded kind of necessity. This *external* or *objective necessity*, of course, is *explanatorily* dependent upon your own beliefs: it is *because* a certain pattern of beliefs is the case (that p and that if p, then q) that you necessary fall under a certain demand. But this *external* necessity - and here comes the important point - is *not* experienced by your as dependent on any policy you endorse. Rather, it is experienced as a fully external constraint, one beyond of any decision from you. But if this feeling is so salient, how should we accommodate it in the debate at the level of *Meta-rationality*?[163]

As far as *truth* is a semantic property entirely independent of any commitment of yours, a realist tale about deliberation seems to work much better than a plan-laden account. Of course, this is not uncontroversial. But the point is that if it were so, realists could argue that your second-order normative beliefs would be better conceptualized as straight, full-blooded beliefs, i.e. as beliefs tracking an independent property (the truth of p and the truth of if p, then q) that establishes the truth of a normative property of yours (that you ought to believe that q, if such set of beliefs is the case). Now, where is the explanation offered by expressivism?[164]

Although to some people realism sounds as an extremely plausible account of our feelings about deliberation, I think that expressivists can manage themselves to manoeuvre here. They could argue, for instance, that whether a certain policy (P) is beyond an *existential* kind of commitments has to be assessed not by appealing to a semantic property (truth) but rather by asking whether P is supported by *higher-order norms of acceptance*. *Higher-order norms of acceptance* are norms governing the acceptance of other norms [Gibbard, A. 1990. p. 168]. In

---

[163] The feeling is platitudinous, as Gibbard himself recognizes. In speaking about what we *mean,* Gibbard notes: "When a person calls something rational, he seems to be doing more than simply expressing his own acceptance of a system of norms (…) He claims to speak with authority; he claims to recognize and report something that it is true independently of what he himself happens to accept or reject" Gibbard, A. 1990. p. 153. In the same paragraph Gibbard makes clear that no error-theoretic solution should be acceptable about this core feature: "He [the person] claims the backing of considerations that, in some sense, 'compel acceptance' of what he is saying. Perhaps he is wrong, but that is the claim he is making. Any account (…) that ignores this claim must be defective. It may well capture all the speaker is saying without illusion, but it will not capture all that he in fact is claiming". See Cuneo, T. 2006 for an interesting treatment of this constraint for expressivism

[164] My question assumes a certain reading of cognitivism at the level of *Meta-rationality*. Jackson, for instance, writes: "If someone believes that P and believes that if P then Q, then they have the property of being such that they ought to believe that Q. Jackson, F. 2001. p. 102

Gibbard's sense, *coherence* and *informedness* are higher-order norms of acceptance. They demand of us to accept P only when P is not taken to be inconsistent with other norms (*coherence*) and when the acceptance of P is minimally informed (*informedness*). In Gibbard's sense, when we accept P and P is not against any higher-order norm of acceptance (as the ones just referred) we could take P as a *permitted* policy. In he same way, when P is *required* by a higher-norm of acceptance P is a *demanded* policy. In Gibbard's terms, basic rules of rationality (and basic rules of morality) should be understood as *required policies* - as policies whose acceptance is somehow constitutive of achieving a certain degree of social coordination and personal self-management. Because of this distinction, we can separate epistemic and evaluative policies from mere existential commitments [Gibbard, A. 1990. p. 169. Blackburn, S. 1992. p. 346][165].

But let me slow down here. Assuming that the account described above is initially plausible, it proves at the very least that we have a theoretical basis to distinguish between policies and existential commitments. But how could we use it to explain the difference between a policy and a simple existential commitment as experienced by yourself, from your first-personal standpoint? To answer this question, let me start by suggesting that higher-order norms of acceptance are supported by a set of perceptual, inferential, and emotional dispositions. But if I am on the right track here, these dispositions would give rise to the very same set of responses we usually cite to illustrate our own feeling of objectivity. Thus, if the problem about the phenomenology of objectivity is equivalent to explain, for example, why an agent cannot fully accept a prohibition against bullfighting as authoritative, while believing at the same time something like

 (1) If I had thought differently bullfighting would have been right

expressivism could plausibly accommodate the *perceived objectivity* of our basic rules of rationality by using second-order norms of acceptance. In these terms, an agent could not

---

[165] [165] A plan-laden account implies, of course, the very same requirements and prohibitions than a truth-functional approach to necessity. Gibbard writes: "Decided states thus work in a way isomorphic to the working of truth-conditions. That is to say, the structure of *allowing* decided states or not is the same as the structure of being *true* or not under determinate conditions. Just as a disjunction can be treated as a truth-functional connective, we can treat it, in a like manner, as an '*allowing*-functional' connective". Gibbard, A. 2003. p. 46

properly respond to a basic rule of rationality (understood as an epistemic policy) if she believed something like

    (1') If I had endorsed another epistemic policy, not believing that q while still believing that p and that if p then q would have been right

And this is so because her higher-order norms of acceptance would enforce a set of dispositions, which would include a disposition not to entertain the content expressed by (1') when faced with a requirement such as BC– because such belief would minimize our chances to obey BC, lowering our opportunities to coordinate with other people and with ourselves. Therefore no recourse to truth or representation is required to explain the type of objectivity we say to perceive in deliberating. First-order and second-order levels of acceptance would be sufficient to explain our externally grounded feeling of objectivity.

If this argument were right, nevertheless, it would seem as if expressivism is able to accommodate the objectivity we literally perceive in first-personal contexts of deliberation. Moreover, it would seem as if expressivism accommodates such feeling without assuming an error-theoretic account. It is not as if expressivist explained how we perceive our deliberative locus by offering a debunking account of our phenomenology. They would rather make sense of our phenomenology by offering a wider explanatory context where the rationale and function of such phenomenology is better grasped. They do not need to deny, in sum, that we are essentially mistaken by believing a normative content. As far as their favoured account of belief departs from our common sense of belief, they are not licensing mistake or illusion at all.

In sum, when the explanation offered by expressivism is in place, there is no need to recourse to a substantive notion of belief. We are not tracking any sort of property by believing the content of BC $_{\text{plan-laden beliefs}}$. We are simply voicing a certain commitment along a propositional form (*quasi-believing* it), whose basic rationale is to coordinate our personal performances in order to live in a public sphere. But if this were the case, expressivism about rationality is not debunked by an appeal to phenomenology.

5. Conclusions

It is widely assumed that the debates about the *objectivity* of our moral opinions have evolved in recent years [Railton, P. 1996. p. 49. Rosen, G. 1998. p. 386]. The present work has tried to relocate the debate about *moral* objectivity on a new, supposedly less controversial, ground. Accordingly, I have assumed all along this work that the kind of phenomena which is the focus of *Theory of Reasons* and *Theory of Rationality* could offer a better standpoint to answer some questions about morality. If morality is simply another normative domain constructed around the concept of *reason*, why not start by dilucidating the kind of psychological capacities that we exempify in responding to reasons of *rationality* to know the kind of psychological state that underlies our moral opinions?

But this route has proved quite problematic. Even if our responses to reasons in the context of rationality were mediated by beliefs, i.e. by *second-order beliefs* in Pettit's sense, we would need an additional argument to support *cognitivism* at the level of *Meta-rationality*. And this is so because some recent *expressivistic* approaches accept that many of the features and regularities that define the stance that we occupy when we respond to deontic normative facts (ought-facts) can be properly described as *non-descriptive beliefs*. *Non-descriptive beliefs* do not purport to represent facts, as ordinary beliefs do. Instead, *non-descriptive beliefs* play a central role in guiding our action through reasoned deliberation [Horgan, T. Timmons, M. 2006. p. 231-233] and in helping to coordinate our actions at an intrapersonal and interpersonal level [Gibbard, A. 2003. p. 40-49].

If the expressivistic recourse to non-descriptive beliefs is sound, we should work further to precise the type of belief that would structure our responses to reasons of rationality. In doing so, we should turn our attention to core notions such as *expression*, *representation*, or *assertion,* and determine their basic functions and the psychological processes that could enable their normal functioning. Only by studying these core concepts we could ascertain to what extent some psychological questions about reasons and rationality are relevant to determine the status of our judgments of reasons and rationality. And by extension, only by studying these core concepts we could achieve a far reached consequence about the status of our judgments of rationality. Thus, the shift toward issues of *Meta-rationality* could be

improved by adding a general concern about issues of *expression*, *social coordination*, and the mediating role played by our *normative assertions* with respect to these general phenomena.

CONCLUSIONES GENERALES

Este trabajo se ha centrado en la defensa de la siguiente tesis:

(M) Si queremos hacer *Meta-ética* de modo provechoso entonces debemos atender a ciertas cuestiones de segundo nivel sobre nuestras razones y nuestra racionalidad.

Cuatro puntos merecen destacarse en relación con M.

En primer lugar, M es una tesis *metodológica*. M sugiere un posible modo de atajar cierto *impasse* entre cognitivistas y no-cognitivistas sin comprometerse con una tesis propiamente meta-ética. M simplemente defiende que el análisis de nuestra capacidad para responder a ciertas demandas de racionalidad podría facilitar una perspectiva adicional para hacer *Meta-ética.*

Aunque M no sea una tesis propiamente axiológica, M presupone una cierta imagen sobre el contenido o el *foco* de la *Meta-ética.* Y esto es lo segundo a tener en cuenta sobre M. Según M, una serie de cuestiones ligadas a la *fundamentación* de nuestras opiniones morales podrían verse reducidas a una única cuestión, una sobre el tipo de estado psicológico que expresamos al proferir un determinado enunciado moral.

En tercer lugar, M hace suyo un compromiso meta-normativo básico. M se compromete con cierta versión de lo que se viene denominando *racionalismo moral.* Aunque *racionalismo* es una etiqueta bastante difusa, el sentido que M hace suyo tiene la fuerza de ciertas verdades conceptuales. Supongamos, para ilustrar el estatuto conceptual ligado al *racionalismo*, que durante una discusión con B, A profiere la siguiente sentencia

(1) Debes evitar que tus hijos digan mentiras

B, un individuo ciertamente peculiar, pregunta por qué debe evitar que sus hijos digan mentiras. A podría responder a B:

(2) Porque es moralmente incorrecto decir mentiras

(3) Porque simplemente tienes que evitar que ellos hagan eso!

(4) Porque decir mentiras les podría acarrear algún castigo

(2), (3) y (4) deberían producirnos cierta perplejidad, cierta extrañeza. Atendiendo a esa extrañeza, sugiero, podríamos entender en qué difiere la fuerza *moral* asociada a la proferencia de (1) de aquella prescriptividad ligada a otras oraciones convencionalmente proferidas con la intención de influir en la conducta de otros agentes.

En mi trabajo no asumo una tesis general para explicar el tono específicamente moral ligado a la proferencia de una determinada oración P. Parece claro, no obstante, que el tono moral de P depende del tipo de razones a las que podríamos aludir para justificar el acto demandado por P *sin provocar extrañeza en un hablante competente*. Identificar el tono moral de (1), por tanto, supone indagar en la extrañeza que nos producen respuestas como (2), (3), y (4). Los rasgos básicos ligados a esas respuestas conformarían,, si estoy en lo correcto, una especie de negativo a partir del cual podríamos delimitar los contornos de aquellas razones que justifican el contenido proferido por (1) *en sentido moral*.

Las razones que emergen tras aplicar este procedimiento son peculiares. De entrada, parece que esas razones son válidas con independencia de que cierta sanción sea aplicable en el contexto donde (1) se profiere – algo que (4) cancela de modo explícito. Igualmente, las razones a favor o en contra del contenido de (1) son razones *morales* sólo si pueden ser válidas con independencia del estatuto moral de esa acción. La propiedad de *ser incorrecta* no podría facilitar por sí mismo una razón para no realizar una determinada acción - algo que (2) y (3) niegan.

Cuando aplicamos la metáfora del negativo a lo que acabo de decir apreciamos un hecho simple: el contenido de nuestras preferencias morales se justifica con independencia de los deseos de aquellos a los que se interpela a través de la proferencia de ese contenido, de las sanciones institucionales que resultan de esa proferencia y del hecho de que la acción demandada sea incorrecta. *Correcto* e *incorrecto* siempre requieren, al menos en el sentido que he favorecido en esta investigación, una apelación a razones, a ciertos rasgos generalmente valiosos que podrían ser respetados o fomentados al actuar de cierto modo. Dicho esto, una buena respuesta por parte de A a las dudas de B sería

(5) Porque decir mentiras no promueve la confianza

(5) apunta a una razón incondicional, no dependiente de un conjunto de sanciones y relacionada de modo directo con la promoción o el respeto de un valor – *confianza* – esencial para la vertebración de cualquier agregado de individuos. (5) es el reverso de aquellos rasgos que nos resultaban peculiares o extraños en (2), (3) y (4). Como tal, (5) facilita una razón genuinamente *moral* que sustenta la autoridad de una proferencia *moral* – (1). Lo anterior esquematiza el sentido de racionalismo que tengo en mente para sustentar M.

Finalmente, M asume que aquellos *hechos normativos* a los que respondemos al comportarnos *moralmente* pueden identificarse con aquellos *hechos normativos* a los que respondemos cuando nos comportamos *racionalmente*. Esta equivalencia resulta esencial para la suerte de M. Sólo defendiéndola podemos afirmar – como M hace – que los procesos psicológicos que conforman nuestra racionalidad pueden funcionar como indicadores o marcas fiables de aquellos procesos psicológicos que ejemplificamos al comportarnos moralmente. Si en los dos contextos (*moralidad* y *racionalidad*) rastreamos y respondemos al mismo tipo de *razones,* M es inicialmente plausible como consejo metodológico para determinar el contenido psicológico de nuestras opiniones morales.

En lo que sigue voy a ocuparme de aquellas conclusiones que podrían derivarse de nuestro compromiso con M. Las consecuencias que conlleva aceptar M se estructuran en torno a tres grandes dominios. (i) Un dominio *meta-metodológico*. (ii) Un dominio puramente *metodológico*. (iii) Un domino *sustantivo* ligado al estatuto del *racionalismo* y a las consecuencias que un análisis de segundo nivel sobre nuestras capacidades racionales podría acarrear para la comprensión del estatuto de nuestros juicios morales. En lo sucesivo prestaré atención a estas tres dimensiones, comenzando por aquellas que tienen que ver con el estatuto de la *Meta-ética* como disciplina.

Conclusiones meta-metodológicas

La primera parte de mi investigación se ha estructurado en torno a la exposición del ideal metodológico en torno al cual gira la *Meta-ética*. En términos de ese ideal, cuando hacemos *Meta-ética* debemos *acomodar,* dentro del molde suministrado por un conjunto de disciplinas filosóficas, una serie de intuiciones y rasgos formales ligados a nuestra perspectiva moral.

Describir la perspectiva desde la que hacemos *Meta-ética* es tarea compleja. Baste decir que el molde sobre el cual *acomodamos* explicativamente nuestras intuiciones morales se ve limitado por ciertos compromisos meta-filosóficos básicos (*naturalismo*) y que no se conforma atendiendo a un conjunto estático de disciplinas filosóficas. Durante el segundo capítulo de mi tesis he tratado de exponer la evolución de esa perspectiva analítica de segundo nivel. En ese capítulo he señalado cómo la evolución histórica de la *Meta-ética* ha favorecido una oposición entre dos imágenes filosóficas en torno al estatuto de nuestra práctica moral. Allí sugiero que la oposición entre estas dos imágenes se ha difuminado en los últimos años como consecuencia de ciertas tesis filosóficas.

Dejando a un lado la génesis de esta progresiva indiferenciación, en mi tesis he acentuado dos consecuencias meta-metodológicas que emergen cuando entendemos la *Meta-ética* como un proyecto filosófico de *acomodación*. De un lado, parece evidente que si reconociésemos la centralidad del ideal de *acomodación* cierta desazón metodológica podría ser explicitada. Por decirlo de modo breve: si asumiéramos el ideal de *acomodación*, entonces podríamos modelar o explicitar nuestro desasimiento teórico ante la creciente sofisticación que reina en la *Meta-ética* como consecuencia de la citada indiferenciación.

Pero hay otra consecuencia importante que se derivaría de aceptar la centralidad del ideal de *acomodación*. Al asumir ese ideal no sólo seríamos capaces de explicitar un sentimiento de desasimiento teórico previamente inarticulado. Además podríamos facilitar un marco general que minimizaría ese desasimiento. Para ejemplificar esto supongamos que nuestro ideal de *acomodación* funciona como una bisagra entre dos niveles. Por un lado, un primer nivel centrado en un conjunto de intuiciones sobre en nuestra perspectiva moral – *foco*. Por otro, un nivel compuesto por una serie de disciplinas filosóficas – *contenido*. La situación de indefinición a la que vengo aludiendo (*impasse*) podría resolverse de dos maneras distintas según esta imagen general:

A. A través de la inclusión de una nueva instancia explicativa dentro del *contenido* de la *Meta-ética* - o a través de la inclusión de un nuevo compromiso que suplemente el rol que la convicción naturalista juega en nuestros procesos de acomodación.

B. A través de la ampliación del *foco* o *dominio* sobre el cual la *Meta-ética* emite sus juicios.

En mi investigación me ocupo de dos estrategias basadas en B. La primera señala que el mejor modo de superar el *impasse* pasaría por ampliar el *foco* de interés de la *Meta-ética* de modo que ésta se ocupara de algunas cuestiones que emergen a partir de la consideración de nuestro concepto de *razón*. En la parte intermedia de mi investigación me centro en esta estrategia general, articulándola en torno a dos esquemas argumentativos generales que se aprovechan de los dos sentidos comúnmente ligados a nuestro concepto de *razón* - un sentido explicativo y uno normativo.

La segunda vía para salvar el *impasse* a partir de B la desarrollo en la última parte de mi investigación. Esta ruta no niega la centralidad de nuestro concepto de *razón* a la hora de entender nuestra perspectiva moral o el tipo de autoridad que caracteriza a nuestros juicios morales. La novedad con esta estrategia reside no obstante en que no exige asumir ninguna tesis sustantiva acerca del estatuto de nuestras *razones*. La estrategia que favorezco se concentra en un análisis de segundo nivel centrado en algunas situaciones en las que nuestro concepto de *razón* se aplica de manera consistente. Me ocupo de aquellas situaciones en las que deliberamos para cumplir con una demanda de racionalidad. La alternativa que favorezco propone replicar el mismo tipo de preguntas que caracterizarían a la *Meta-ética* pero en relación con esas situaciones .

En definitiva, al adoptar el ideal de *acomodación* como el principio guía para evaluar las explicaciones facilitadas por nuestras teorías *meta-éticas* dos consecuencias resultan evidentes:

- **(C1)** Si el ideal de *acomodación* fuese asumido como principio guía entonces podríamos *explicitar* la estructura general que subyace a cierto escepticismo sobre la disponibilidad de una explicación unitaria de la naturaleza de nuestra perspectiva moral

- **(C2)** Si el ideal de *acomodación* fuese asumido como principio guía entonces podríamos apelar a ese ideal para modelar y corregir esa situación de indefinición teórica.

Conclusiones metodológicas

El ideal de *acomodación* podría ayudar a modelar el desacuerdo que enmarca gran parte del debate reciente en *Meta-ética*. Poner sobre el tapete la importancia de este ideal para entender ese desacuerdo, no obstante, dice bastante poco sobre cómo deberíamos superarlo. En esta sección trataré de ser más explícito acerca de ese aspecto, esbozando dos conclusiones metodológicas que pueden colegirse del material que he presentado en la primer parte de mi tesis.

La primera conclusión claramente metodológica que podría extraerse de mi investigación es bastante general. Asumiendo que la *Meta-ética* se concentra en responder a una cuestión psicológica, la novedad del enfoque que favorezco en la parte final de la tesis reside fundamentalmente, como ya he apuntado, en el *foco* de indagación psicológica que asocio con la *Meta-ética*. Mientras que es una práctica estándar depositar muchas esperanzas en ciertos análisis que toman las regularidades asociadas con nuestras actitudes básicas (creencia, deseo) como foco de interés para la *Meta-ética*, es bastante menos común trasladar ese interés a ciertas cuestiones relacionadas con aquellos procesos psicológicos que ejemplificamos al responder a ciertos hechos normativos, es decir, a ciertas razones que sustentarían la autoridad de nuestras reglas de racionalidad. Cuando analizamos este tipo de fenómenos en clave psicológica, aún sin variar contenido de la *Meta-ética* logramos ampliar su *foco*. Hacer *Meta-ética* a través de un análisis del estatuto psicológico de estos fenómenos constituye el compromiso metodológico fundamental en mi trabajo.

Asumamos que un enfoque psicológico centrado en nuestra capacidad para responder a ciertos hechos normativos puede ayudar a superar el *impasse* en que la *Meta-ética* se encuentra estancada. Aún aceptando que el contenido de este principio metodológico es más o menos directo, su alcance en relación con aquellas disciplinas que conforman la *Meta-ética* en sentido tradicional debería ser precisado. Y aquí adopto una postura moderadamente

maximalista. Esencialmente defiendo que podríamos comenzar a esbozar una teoría *meta-ética* prestando atención de modo exclusivo al aspecto psicológico que ejemplificamos cuando respondemos a ciertos hechos normativos. Pero obviamente esto no supone negar que sería necesario decir muchas más cosas sobre esta capacidad y sobre cómo se conecta con nuestros juicios morales. En cualquier caso hacer *Meta-racionalidad* sería, como mínimo, un buen modo de iniciar el debate en *Meta-ética*.

Dos postulados metodológicas sustentarían, por tanto, mi investigación

- **(C3)** Un buen modo de superar el *impasse* que caracteriza a la *Meta-ética* consistiría en analizar, en clave psicológica, el tipo de capacidades que ejemplificamos en aquellas situaciones en las que respondemos a un hecho normativo conformado en torno al concepto de *razón*

- **(C4)** El tipo de análisis favorecido por (C3) constituiría una base *suficiente* para establecer un conjunto de tesis *ontológicas*, *semánticas* y *epistemológicas* sobre nuestros juicios morales

C3 es equivalente a M, la tesis básica de esta investigación

> **(M)** Si quisiéramos hacer *Meta-ética* de modo provechoso entonces deberíamos atender a ciertas cuestiones de segundo nivel sobre nuestras razones y nuestra racionalidad.

Defender la solidez de C3 equivale, por tanto, a apuntalar M. Para apuntalar C3 debemos asegurar varios supuestos de índole más sustantiva. C3 implica además una serie de consecuencias sobre el estatuto de nuestra racionalidad, sobre su autoridad y sobre el tipo de procesos mentales que ejemplificamos al comportarnos racionalmente. Defender C3 equivale, por tanto, a determinar si aceptamos estas consecuencias o no. En lo que sigue paso revista a los supuestos que soportan C3 y a las consecuencias que se derivarían de aceptar C3.

Conclusiones sustantivas

Empecemos por tres tesis que deberían asumirse para sustentar el ideal metodológico favorecido por C3.

Como he señalado al principio, en la medida en que C3 trata de reorientar el *foco* de la *Meta-ética* atendiendo a ciertos fenómenos de segundo nivel localizados en el ámbito de nuestra racionalidad, lo que algunos filósofos denominan como *racionalismo* debería ser un componente esencial en la reubicación metodológica que C3 defiende. El *racionalismo* que favorezco a lo largo de mi investigación no se compromete, en principio, con ninguna concepción sustantiva acerca del estatuto de nuestras razones. Como he sugerido al principio se presenta, más bien, como una tesis centrada en el tipo de explicación que deberíamos proveer cuando alguien cuestiona la normatividad de una determinada obligación. En lo que sigue entenderé que lo que muchos denominan *racionalismo* se puede reducir al siguiente principio:

> **(NR)** Necesariamente, si A *debe* φ entonces A tiene una *razón* (concluyente) a favor de φ

El *racionalismo moral* es una consecuencia *local* de NR. El *racionalismo moral* es verdadero, por tanto, sólo si NR es verdadero. Adicionalmente, si queremos establecer el racionalismo moral y usarlo para sustentar C3 (o M) debemos establecer una equivalencia más fina entre el tipo de *hechos normativos* a los que respondemos al comportarnos *racionalmente* y el tipo de *hecho normativos* a los que respondemos al comportarnos *moralmente*. En los términos favorecidos por NR sería conveniente precisar en qué sentido aquellas *razones* a las que respondemos al actuar *moralmente* son similares a aquellas *razones* a las que respondemos cuando actuamos *racionalmente*.

En mi trabajo he tratado de mostrar que tenemos una base más que suficiente para presuponer cierta equivalencia entre estos dos tipos de *hechos normativos*, es decir, entre aquellos hechos (normativos) que gobiernan el dominio propio de nuestra racionalidad y aquellos hechos (normativos) que determinan nuestras obligaciones morales. Para ilustrar esta equivalencia he asumido que ambos tipos de *hechos* comparten tanto un rasgo de tipo

estructural como un aspecto de índole más sustantiva. Comencemos por una *similitud de tipo estructural*. Supongamos que afirmamos

(1) A debe ɸ

Al proferir (1) estamos asumiendo que cierto hecho es el caso - con independencia de que la *obligación* expresada por (1) se derive de un sistema de reglas morales o de un conjunto de principios derivados de ciertos patrones de interacción funcional establecidos entre nuestras actitudes. En cierto modo, si (1) entonces

(B) N favorece ɸ

B es un hecho normativo constituido alrededor del concepto de *razón*. B se conforma en torno a una relación básica de *favorecimiento,* una relación que se establece entre un *hecho* o conjunto de hechos no-normativos (N) y una *acción* o *actitud* (ɸ) - creencia, sentimiento o intención - que es favorecida por N. La determinación normativa o fuerza expresada por esta relación se conforma, explicativamente al menos, a partir N. Cuando la normatividad de (1) se constituye a partir de B decimos que N es la *razón* por la que A debe hacer ɸ. Explicamos la normatividad de (1) o su fuerza en virtud de un concepto normativo más básico, el concepto de *razón*. Una razón favorece una determinada acción o actitud en virtud del tipo de *valor* envuelto en la formación de esa acción o actitud. En cierto sentido, la fuerza de (1) puede explicarse mediante una relación normativa básica en clave de *favorecimiento* por el simple hecho de que cierto tipo de valor es ejemplificado si ɸ es el caso - *ejemplificado* es usado aquí de modo liberal, por supuesto, pudiendo sustituirse por relaciones más complejas como *promoción* o *respeto*.

Además de compartir esta estructura general, he defendido que los hechos normativos de tipo *moral* comparten un rasgo adicional con respecto a aquellos hechos que gobiernan nuestras actitudes en el plano racional. Según esta similitud adicional, ambos tipos de hechos canalizan nuestras actitudes o nuestro comportamiento a través del mismo tipo de *autoridad*. A pesar de que la aplicación de ciertos principios de racionalidad depende de qué actitudes estemos ejemplificando en una determinada situación, el hecho de que debamos formar una actitud como resultado de este hecho contingente *no* depende de ningún motivo o deseo por

comportarse de forma racional. En la medida en que creemos que *p* y creemos que si *p* entonces *q* debemos creer que *q* - independientemente de nuestros intereses o deseos. Igualmente, en aquellos casos en los que B está pasándolo mal debemos atender a B con independencia de que la ayuda que brindemos a B pueda facilitar algunos de nuestros proyectos o satisfacer algunas de nuestras preferencias. Las dos obligaciones expresan, por decirlo de modo familiar, obligaciones categóricas sobre nuestras actitudes, obligaciones cuya fuerza o prescriptividad es enteramente independiente de nuestros intereses o deseos - aún cuando esas obligaciones requieren, a la hora de aplicarse, que el agente ejemplifique ciertas actitudes.

Pero a pesar de la plausibilidad del *racionalismo* y de las similitudes que acabo de apuntar entre aquellos hechos normativos de tipo moral y aquellos hechos normativos de índole racional, asumir C3 requiere hacer frente un problema incómodo. El problema al que me voy a referir fue inicialmente apuntado por Michael Bratman. Bratman lo introdujo al hilo de su crítica a la teoría davidsoniana de la acción. Lo que Bratman denomina allí *bootstrapping* no busca primariamente demostrar la falsedad de NR. En mi trabajo, sin embargo, apelo a casos de *bootstrapping* para sustentar una crítica sobre NR.

Supongamos, para ilustrar cómo funciona esta crítica sobre NR, que decido salir de copas esta noche a pesar de que mañana tengo que defender esta tesis. Supongamos que hay bastantes razones que favorecen que me quede en casa y ninguna a favor de salir de copas hasta muy tarde. Tras deliberar un momento, no obstante, decido ir al centro a tomar unas cervezas con los amigos. Asumamos que para llegar al centro antes de que mis amigos decidan irse a otra zona de la cuidad tengo que coger un taxi. Es tarde y los autobuses, al menos en Granada, no funcionan por la noche. Una vez llegados a este punto, parece como si *debiera* coger un taxi, es decir, como si estuviera obligado a coger un taxi en virtud de cierto principio de *racionalidad*. Parece, en definitiva, como si cierta regla básica de racionalidad (*principio de racionalidad instrumental* - PRI) se aplicase en la coyuntura que ocupo, obligándome a formar una determinada intención en virtud de una intención previa y cierta creencia instrumental. Si soy racional, por decirlo de algún modo, estoy obligado a coger un taxi en caso de que quiera ir al centro a tomar una cerveza con mis amigos y creo que el único modo de llegar hasta allí es cogiendo un taxi.

Pero ahora viene el punto problemático. Según NR, podríamos explicar la necesidad o prescriptividad ligada a PRI en esta situación si hiciésemos referencia a ciertos rasgos

valiosos que se derivarían de mi intención de coger un taxi en presencia de mi intención de tomar una cerveza con mis amigos (y mi creencia de que un coger un taxi es un medio necesario pasada cierta hora para ir al centro). Podríamos explicar, en suma, la normatividad ligada a aquella opción demandada por PRI en términos de ciertas *razones* que favorecerían que cogiera un taxi. Así, si NR fuese correcto, el mero hecho de que mis estados mentales se estructurasen en torno a un cierto patrón instrumental sería suficiente para justificar cierta intención. No obstante, resulta evidente que *no* tengo ninguna razón que favorezca coger un taxi para ir al centro. Ningún valor sería actualizado si cogiera el taxi - y ciertos valores negativos serían actualizados lo cogiera. De este modo, en la medida en que ninguna razón justifica la opción requerida por PRI, NR es falso en este caso concreto. En algunas situaciones estamos obligados a formar una determinada actitud sin que podamos justificar nuestra obligación en virtud de cómo ciertos rasgos valioso favorecerían la formación de esa actitud. Nuestras actitudes, por decirlo de modo intuitivo, no pueden facilitarnos una razón a favor de cierta actitud simplemente en virtud de como ellas ejemplifican ciertos patrones funcionales.

La consecuencia de este tipo de casos debe quedar bien clara: puesto que la plausibilidad del consejo metodológico facilitado por C3 requiere la verdad de NR (y NR, es verdadero sólo si la normatividad de *todas* nuestras demandas puede explicarse en clave de razones), debemos colegir que C3 es implausible en virtud de casos como el que acabo de esbozar. C3sería incapaz, por tanto, de sustentar un reubicación del foco sobre el cual gira la *Meta-ética*. Es simplemente falso que siempre que un debe se nos aplica podamos justificar su fuerza apelando a razones.

¿Cómo de grave es esta recusación de C3? No definitiva, ciertamente. Si nos centramos en la forma lógica que subyace a nuestros principios básicos de racionalidad podemos minimizar el efecto combinado que tiene la acusación de *bootstrapping* sobre NR y C3. Como todos sabemos, esta línea de respuesta ha sido favorecida recientemente por John Broome. Según Broome, nuestros principios básicos de racionalidad deberían interpretarse como esquemas lógicos en los que cierto operador deóntico básico (*debe*) gobierna el condicional que normalmente subyace a esos principios de racionalidad. Nuestros principios de racionalidad no deberían entenderse, por tanto, como esquemas lógicos en los que el citado operador gobierna únicamente el consecuente del condicional. Así, cuando aplicamos PRI sobre A en virtud de sus intenciones y creencias, no aseveramos que si A intenta *p* y A cree que *q* es un

medio necesario para *p* entonces A *debe* intentar *q.* Es decir, al evaluar a alguien en términos de PRI no presuponemos que

(NS) $(p, p \rightarrow q) \rightarrow D q$

sino que localizamos el alcance del operador deóntico (D) sobre el condicional conformado por aquellas actitudes que posibilitarían la aplicación de PRI. La siguiente formula expresa lo que acabo de decir

(WS) $D ((p, p \rightarrow q) \rightarrow q)$

Cuando la forma lógica que subyace a PRI se entiende en términos de WS decimos que el operador deóntico expresado por PRI tiene un *alcance amplio* (*wide-scope*). Si la forma lógica se entiende, en cambio, en términos de NS entonces asumimos que el operador deóntico tiene un alcance estrecho (*narrow-scope*).

Una interpretación *amplia* de las fórmulas condicionales subyacentes a cualquier principio de racionalidad acarrea dos consecuencias inmediatas que detallo en mi trabajo. La primera consecuencia es que no podemos colegir una obligación incondicional - Dq - a partir de (WS) y $(p, p \rightarrow q)$. La segunda consecuencia, de cariz más positivo, precisa el tipo de obligación que podría derivarse de una interpretación *amplia* de nuestras reglas básicas de racionalidad. Si Broome está en lo correcto, cuando PRI se entiende como WS prescribe, una cierta combinación de actitudes está prohibida para aquellos agentes a los que podemos aplicar PRI

(WS **banned)** $\neg D ((p, p \rightarrow q) \wedge \neg q)$

A partir de la combinación de actitudes excluida por WS, podríamos derivar una obligación *condicional* si negásemos la formula conjuntiva que aparece arriba

(WS*) $D (q \vee \neg (p, p \rightarrow q))$

WS* requiere, llegado t_2, que A intente que *q* sea el caso o que A abandone una de las actitudes que motivo la aplicación de PRI (es decir, que abandone la intención de tomar una

cerveza con los amigos o la creencia instrumental) en aquellas situaciones en las que, en t_1, A intentó que p fuese el caso y creyó que si p entonces q. Según esto, nuestros principios de racionalidad expresarían obligaciones condicionales con la misma estructura que WS*.

Cuando estos dos puntos ligados al alcance de los operadores deónticos se asumen, la recusación conjunta de NR y C3 a partir de ciertos casos de *bootstrapping* se desactiva. Si WS expresase el modo correcto de entender nuestras reglas de racionalidad, podríamos justificar la prescriptividad de *todas* aquellas situaciones en las que nuestras reglas de racionalidad se aplican apelando a razones (como NR sugiere). Y podríamos hacerlo sin presuponer que el mero hecho de ejemplificar ciertos patrones funcionales entre nuestros estados mentales facilita una razón para formar una determinada actitud.

Volvamos al caso anterior – el caso en el que debo coger un taxi sin poder citar ninguna razón a favor de esa opción. Básicamente, cuando distinguimos entre WS y NS no tenemos por qué colegir que en aquellos casos en los que A intenta que p y A cree que q es necesario para que p sea el caso A deba necesariamente intentar que q sea el caso. Si WS está en lo cierto, cuando A intenta que p sea el caso y A cree que q es un medio necesario para intentar que p, A está obligado a intentar que q sea el caso o a revisar su intención o su creencia instrumental. A no esta obligado en ningún caso a intentar que q sea el caso. Por tanto, A podría explicar (y justificar) su seguimiento de PRI apelando a una *razón condicional*. Esta razón favorecería, en el caso concreto de A, una intención a favor de q en caso de que A intentase que p y creyese que q es un medio necesario para p o una revisión de sus actitudes en caso de A no formase una intención a favor de q. Si la normatividad de PRI se pudiera justificar apelando a este tipo de razones - *razones condicionales* – entonces NR sería correcto. Y, por supuesto, C3 podría seguir conformando un ideal metodológico plausible.

Dos conclusiones sustantivas deben presuponerse, por tanto, si C3 es aceptado como consejo metodológico:

- **(C5)** NR es verdadero, luego la acusación de *bootstrapping* es infundada: La normatividad asociada a nuestros principios de racionalidad puede *siempre* explicarse como NR sugiere sin que esto presuponga que nuestros estados mentales justifican, por sí mismos, la formación de una determinada actitud

- **(C6)** Aquellas hechos normativos a los que respondemos cuando somos racionales pueden conceptualizarse - tanto desde un punto de vista estructural como desde uno más sustantivo - de modo análogo a aquellos hechos normativos a los que respondemos cuando nos comportamos moralmente.

Además de las dos conclusiones sustantivas apuntadas arriba, en este trabajo he defendido que el proceso a través de cual establecemos C5 lleva aparejado una tesis adicional. Esta tesis está centrada en el tipo de estado mental que posibilita nuestra respuesta ante ciertos requerimientos de racionalidad:

- **(C7)** C5 es plausible porque nuestra capacidad para responder a ciertos hechos epistémicos sobre nuestras actitudes se estructura en torno a capacidades cognitivas de segundo nivel – *creencias normativas de segundo nivel* sobre nuestras propias actitudes.

La defensa de C7 ha ocupado un lugar destacado en mi investigación. Mi defensa viene motivada a propósito de una crítica de Niko Kolodny sobre C5. Kolodny defiende que no podemos interpretar de modo *amplio* la forma lógica subyacente a nuestros principios de racionalidad. Según Kolodny, si prestásemos atención al foco de nuestros procesos de revisión de actitudes no podríamos interpretar el alcance del operador deóntico en nuestras regla de racionalidad según un esquema *amplio*. En la medida en que una interpretación *amplia* de nuestras reglas de racionalidad (WS) es necesaria para asegurar la tesis racionalista (NR), si WS es falsa entonces NR es falsa. Pero concluir que NR es falsa es equivalente a negar C3. Y esto, como anuncié, es equivalente a negar M, la tesis central de este trabajo. Según lo que acabo de decir, la defensa de WS constituye un punto esencial para soportar la tesis central de este trabajo. En lo que sigue explico cómo podría rechazarse WS. Después de eso explico cómo podría defenderse WS.

Supongamos que respetar *de modo autónomo* una regla de racionalidad exige ser capaz de razonar a partir de aquellos estados mentales que posibilitan la aplicación de esa regla de racionalidad. Supongamos, adicionalmente, que razonar a partir de nuestros estados mentales requiere tomar el *contenido* de esos estados mentales como el *foco* de aquellos procesos que estructuran nuestro razonamiento. K expresa estas dos suposiciones

(K') A forma de modo autónomo una actitud (ϕ) como consecuencia de la aplicación de un determinado principio de racionalidad (P) sólo si el *contenido* de aquellas actitudes que posibilitan la aplicación de P justificaría el *contenido* de ϕ.

Kolodny apela a K' para rechazar la solución de Broome para atajar los casos problemáticos de *bootstrapping*. Consideremos PRI de nuevo. Según la interpretación *amplia* de PRI, si intentamos en t_1 que *p* sea el caso y creemos que *q* es un medio necesario para que *p* sea el caso, estos estados mentales nos sitúan bajo el siguiente requerimiento disyuntivo en t_2

(WS*) D (q $\vee \neg$ (p, p $\rightarrow$ q))

Según Kolodny, si WS* constituyera una interpretación plausible de PRI entonces deberíamos poder cumplir con lo requerido por WS* a través de un proceso de razonamiento que respetara K'. En aquellos casos en los que PRI se nos aplica, deberíamos ser capaces de satisfacer cada uno de los dos disjuntos contenidos en WS* a través de un proceso de razonamiento sustentado en el *contenido* de aquellos estados mentales que posibilitan la aplicación de PRI. En términos formales, Kolodny señala que deberíamos satisfacer la siguiente condición si PRI fuese entendido al modo *amplio* propugnado por Broome

(WS* + K') D (q $\vee \neg$ (p, p $\rightarrow$ q)) $\rightarrow$ [podríamos razonar desde el *contenido* de (p, p $\rightarrow$ q) hasta q $\wedge$ podríamos razonar desde el *contenido* de $\neg$q hasta $\neg$ (p, p $\rightarrow$ q)]

Kolodny señala que no podemos cumplir con lo requerido por PRI según una interpretación amplia de ese principio de racionalidad y respetar, a la vez, la condición general fijada por K'. Esto es evidente cuando consideramos el segundo conjunto contenido en el consecuente del condicional incluido en (WS* + K'). Ahí se nos conmina a rechazar cualquiera de las actitudes que motivaron la aplicación de PRI a partir de nuestra negativa a formar la intención de que *q* sea el caso. Pero una pregunta se impone: ¿cómo podemos razonar desde la no formación de una intención ($\neg$q) - algo que sencillamente no es un estado mental - hasta el

rechazo de alguna de nuestras actitudes (¬(p, p →q))? ¿Qué tipo de contenido sustentaría este proceso de razonamiento si ni tan siquiera contamos con el *locus* psicológico de cualquier contenido proposicional, esto es, una *actitud* proposicional?

Kolodny no se centra únicamente en PRI para ilustrar sus críticas. Si asumimos que algunas veces nuestras reglas de racionalidad se formulan en términos explícitamente normativos podemos ilustrar su rechazo de WS de un modo diferente. Supongamos que uno de los principios de racionalidad más básicos demanda no formar φ (una intención, por ejemplo) si creemos que no tenemos razones que sustentan φ. En los términos favorecidos por WS, esta regla de racionalidad debería interpretarse según el esquema que presento abajo

(B -) D (B¬Iq →¬ Iq)

(B-) dice que debemos renunciar a formar una determinada intención (Iq) en aquellas situaciones en las que creemos que no hay razones que justifiquen la formación de esa intención. Como he señalado antes, cualquier demanda interpretada según WS puede ser formulada en términos disyuntivos. En el caso de B- esto equivale a

(B-*) D (¬ Iq ∨ BIq)

Cuando aplicamos K' sobre (B-*) obtenemos que

(B-* + K') D (¬ Iq ∨ BIq) → [podríamos razonar desde el *contenido* de nuestra creencia de que no hay razones para intentar que *q* sea el caso (B¬Iq) hasta ¬Iq ∧ podríamos razonar desde el *contenido* de nuestra intención de que *q* sea el caso (Iq) hasta una creencia centrada en las razones que sustentan q - BIq]

En este caso sí tenemos dos actitudes proposicionales desde las cuales razonar a la hora de cumplir con lo que es demandado por (B-*). Sin embargo, una de las rutas ejemplificadas arriba es problemática. En particular, el segundo conjunto contenido en el consecuente no puede ejemplificar un proceso razonable de revisión. ¿Cómo podemos revisar nuestra creencia acerca de las razones que sustentan una determinada intención – creyendo que tenemos razones para intentar *q* - a partir de la mera formación de una intención centrada en

q? ¿Cómo puede constituir la mera formación de una intención una base suficiente para revisar nuestras creencias acerca de la justificación de nuestras intenciones?

Como señale antes, la relevancia de estos argumentos contra WS es patente. Si no podemos cumplir con lo demandado por PRI o B- en sentido *amplio* a través de un proceso de razonamiento que respete K', entonces WS* no constituye una interpretación plausible de PRI o de B-. El operador deóntico expresado por nuestras reglas básicas de racionalidad tiene, por tanto, un alcance estrecho. PRI debería interpretarse como

(NS*) $(p, p \rightarrow q) \rightarrow Dq$

Y lo mismo se aplicaría a B-.

Como apunté antes, la propiedad más importante de NS* es que a partir de $(p, p \rightarrow q)$ y NS* podemos derivar una obligación incondicional (Dq). Esto, por supuesto, es fatal para quien defiende NR. Puesto que existen al menos dos principios de racionalidad (PRI y B-) que deben interpretarse como NS dice, debemos asumir que WS es falsa. Pero si WS es falsa, entonces NR puede ser criticado apelando de nuevo al *bootstrapping*. En la medida en que la plausibilidad de C3 (o M) requiere la verdad de NR, C3 (o M) debe rechazarse cuando WS se demuestra falsa.

La crítica de Kolodny se aprovecha de que la verdad de WS es esencial para defender, de modo combinado, NR y C3 (o M). De acuerdo con esto, en mi trabajo he defendido WS. Para hacer esto he asumido que K' fija una condición general, pero no una condición sustantiva para interpretar la naturaleza de nuestros procesos de deliberación. K' ejemplifica *algunos* de los procesos que ponemos en funcionamiento al cumplir con nuestras reglas de racionalidad. Pero K' no facilita la *única* vía para cumplir de modo autónomo con nuestros principios de racionalidad – incluso si estos se entienden como WS recomienda. Pero si K' deja espacio para otras rutas deliberativas, la recusación de WS que Kolodny favorece – una sustentada en una lectura unívoca de K – no resulta concluyente. Y el vínculo argumentativo central para este trabajo puede reconstruirse de nuevo, por supuesto. Rechazar una lectura unívoca de K' equivale, por tanto, a rehabilitar WS. En la medida en que WS sustenta la plausibilidad conjunta de WS y C3 (o M), defender C3 (o M) equivale a defender una lectura alternativa de K'.

Para negar la validez incondicional de K' defiendo que somos capaces de cumplir con nuestras reglas de racionalidad (amplias) mediante la formación de *creencias normativas de segundo nivel.* Una *creencia normativa de segundo nivel* es un constructo. En mi investigación he defendido que tal constructo sólo cobra sentido cuando entendemos nuestra capacidad para responder a ciertas razones de modo disyuntivo. Así, aunque algunas veces tratamos de determinar nuestras creencias y nuestras intenciones a través de un proceso de deliberación, en innumerables ocasiones no necesitamos el concurso de ningún acto deliberativo para actuar de manera racional. El buen funcionamiento de nuestras capacidades perceptivas unido a nuestra capacidad para ajustar nuestras creencias de modo automático a la luz de la evidencia disponible nos permite ser (epistémicamente) racionales sin deliberar. Lo mismo se aplica a nuestras actitudes prácticas: en innumerables ocasiones. Formamos intenciones racionales – intenciones que reflejan las razones disponibles en un determinado contexto – sin el concurso de una actividad deliberativa consciente.

A pesar de lo que acabo de señalar, *deliberar* resulta central para nuestra racionalidad. Deliberar es determinar nuestras actitudes (no siempre con éxito) a través de un tipo de actividad atencional centrada en aquellos rasgos que percibimos como valiosos. Cuando deliberamos tratamos de responder a ciertas razones a través de un proceso que va más allá de ciertos automatismos o hábitos. Mediante ese proceso facilitamos que esos rasgos valiosos determinen nuestras emociones, creencias, e intenciones de manera más fiable. Una cuestión esencial para la filosofía práctica es entender cómo se desarrolla este proceso de deliberación ¿Cómo estructuramos nuestra atención a la hora de facilitar nuestra autonomía como agentes a través de nuestra capacidad para deliberar? Aunque éstas preguntas quedan aparentemente muy lejos de las motivaciones de este trabajo, en un momento determinado abogo por una respuesta general, una que niega una lectura unívoca de K'. Aunque nuestra racionalidad y nuestras capacidades deliberativas no siempre se superponen, en aquellos casos en que ambas coinciden sugiero que no resultaría descabellado suponer que es nuestra habilidad para formar *creencias normativas de segundo nivel* nos permite deliberar de modo más racional, respondiendo a aquellas razones que se nos aplican.

 Si estoy en lo cierto este tipo de creencias sobrevienen sobre (i) nuestras *creencias de segundo nivel* - nuestras creencias sobre nuestros propios estados mentales - y sobre (ii) nuestra capacidad para ser sensibles a las consecuencias derivadas al formar ciertos estados mentales con independencia de esos patrones (y aquí incluyo entre esas *consecuencias* las

reacciones de otros agentes a nuestra desviación). A partir de estas dos habilidades, defiendo, somos capaces de formar creencias cuyo contenido puede especificarse haciendo referencia a una demanda impuesta sobre nuestras propias actitudes, esto es, *creencias normativas de segundo nivel.*

Es importante hacer una puntualización general. En mi trabajo no defiendo que respondemos a aquellas razones que se nos aplican en un determinado contexto sólo si somos capaces de formar creencias sobre esas razones. Como he señalado arriba, esto es claramente falso en relación con aquellos casos en los que somos racionales de modo más o menos inconsciente. Pero en casos típicos de deliberación, casos donde tratamos de influir de modo consciente en nuestras actitudes, eso tampoco es verdadero. Y es que podemos motivar cierta función para nuestras creencias normativas de segundo nivel sin presuponer esa tesis. Por ejemplo, podemos sugerir que en algunos escenarios deliberativos podemos orientar nuestros procesos de razonamiento de primer nivel a través de la formación de creencias normativas de segundo nivel. Podemos revisar nuestras creencias o intenciones de manera más sistemática y fiable porque somos capaces de reflexionar sobre las razones que favorecen cierto estado o cierta combinación de estados. Esto no implica, de nuevo, que las creencias normativas de segundo nivel influyan de modo directo en el resultado de esa revisión. Ellas no causan nuestras creencias. Tampoco las justifican. Al contrario; es el contenido de nuestras creencias de primer nivel – el grado de confianza con el que aseveraríamos cierto contenido o la deseablidad que le suponemos – el que determina el resultado y el estatuto de nuestro razonamiento. Por tanto, aunque nuestras creencias normativas no determinan nuestro razonamiento sí lo pueden dirigir, favoreciendo un mejor cumplimiento con nuestras razones. Aunque sutil, la diferencia que trato de motivar resulta esencial para evaluar la plausibilidad de las creencias normativas de segundo nivel.

¿Cómo se relacionan, no obstante, estas creencias con la línea argumentativa general que vengo discutiendo en estas páginas? En mi trabajo defiendo que estas creencias nos permitirían modelar nuestros principios de racionalidad de modo amplio. Puesto que WS es una condición necesaria para que NR sea verdadero - y C3 (o M) requiere NR – estas creencias resultan esencial para sustentar la tesis central de este trabajo. C7 se refiere a esta intuición del siguiente modo

:

- **(C7)** C5 resultaría plausible porque nuestra capacidad para responder a ciertos hechos epistémicos se estructura en torno a capacidades cognitivas de segundo nivel – *creencias normativas de segundo nivel* centradas en ciertas relaciones normativas entre nuestras actitudes.

Para sustentar la plausibilidad de estas creencias de segundo nivel desarrollo un argumento desde tres niveles. El objetivo es explicitar la plausibilidad de este tipo de creencias frente a quienes presuponen que cualquier tipo de estado de segundo nivel podría minimizar la autonomía y la autoridad del resultado de nuestra deliberación.

En primer lugar, argumento a favor de este tipo de creencias normativas a partir de una imagen muy general, una centrada en el estatuto normativo de nuestra racionalidad. Tal y como presento nuestro concepto de *racionalidad*, cualquier aplicación concreta de una *regla de racionalidad* nos permitiría responder a un valor de extraordinaria importancia, un que subsumo a lo largo de mi trabajo bajo el rótulo de *coherencia*. Por el simple hecho de una regla de racionalidad se nos aplica – muchas veces con independencia de que tengamos razones objetivas para cumplir con lo que esa regla prescribe – podemos responder a un *valor ejecutivo*. Un *valor ejecutivo* es un valor del que cualquier valor sustantivo (*goce estético*, *amistad*, *conocimiento*, *placer*, etc.) depende para ser satisfecho o respetado en nuestra conducta diaria. Nuestra habilidad para ser *coherentes* es un valor ejecutivo en ese preciso sentido. Gracias a ese valor respondemos mejor a ciertos valores sustantivos. En cierto sentido nuestra capacidad para ser coherentes podría entenderse como un *bien primario*. Así como la *justicia* es un bien primario para las instituciones o los agregados de agentes, la *coherencia* sería un bien primario para los individuos y su bienestar pleno. No sería completamente implausible asumir, según esto, que nuestra historia evolutiva nos ha equipado con cierta capacidad para rastrear este tipo de valor primario, es decir, con cierta capacidad para formar creencias normativas de segundo nivel, creencias centradas en las demandas impuestas por ciertas reglas básicas de racionalidad. Estas creencias focalizarían nuestra atención sobre un valor primario de vital importancia, la coherencia entre nuestros estados mentales.

En segundo lugar, he argumentado a favor de esas creencias normativas de segundo nivel remarcando su *función* concreta dentro de la economía cognitiva de un agente en un determinado contexto deliberativo. Dicho de otro modo: Más allá de esa pluralidad de valores que nuestra capacidad para formar creencias normativas podría fomentar, me interesa una

cuestión simple: ¿Cómo resultan útiles nuestras creencias normativas en contextos concretos de deliberación? Para responder a esta cuestión recurro a una variedad de explicación funcional, una favorecida recientemente por Philip Pettit. Pettit asume toda lo esbozado en el párrafo anterior. La novedad introducida por Pettit, no obstante, reside en detallar cómo nuestra capacidad para formar creencias normativas de segundo nivel puede influir en nuestra deliberación real, más allá de su contribución general de cara a mejorar nuestra receptividad ante cierto tipo de valor. Según Pettit, nuestras creencias normativas de segundo orden son útiles simplemente en virtud del nivel adicional que nos suministran para testar el estatuto o la justificación de aquellas creencias o intenciones de primer orden que motivan la aplicación de nuestras reglas de racionalidad – por ejemplo, PRI o BC-. Al ser capaces de formar creencias sobre lo que nuestros estados mentales demandan, podemos maximizar nuestras opciones para cumplir de modo adecuado con cada principio de racionalidad. En cierto sentido - aunque podría argumentarse que cumplimos con nuestros principios de racionalidad a través de aquellos procesos de primer nivel que son el foco de K' - formar creencias de segundo nivel podría incrementar nuestras opciones de revisar nuestras actitudes de primer nivel de manera adecuada. Suponer que hemos sido diseñados para formar este tipo de creencias, por tanto, no resulta demasiado descabellado según Pettit. En la medida en que son útiles y su postulación no contraviene los postulados naturalistas básicos, ¿por qué no aceptarlas como parte de nuestro repertorio de habilidades cognitivas?

Evidentemente, motivar la utilidad de las creencias normativas de segundo nivel en casos concretos de deliberación no constituye el último escollo para argumentar en pos de su plausibilidad. En la medida en que la crítica de Kolodny sobre WS se sustenta en principios de racionalidad *estrechos* (principios parcialmente formulados en clave normativa, como BC-) parece sensato aceptar que cualquier defensa de WS frente a Kolodny debe demostrar que nuestros principios estrechos de racionalidad pueden ser interpretados como WS señala. En mi investigación defiendo precisamente eso. Nuestras creencias normativas de segundo nivel se aplican con independencia del tipo de regla racional que se nos aplique. Con independencia del contenido de una regla R, aquellos que defienden la plausibilidad de las creencias normativas de segundo nivel defienden que maximizamos nuestra obediencia respecto a R si somos capaces de formar creencias normativas de segundo nivel sobre lo que R demanda.

. Hasta aquí la justificación del rol central que C7 juega en mi investigación Llegados a este punto una pregunta central parece imponerse: ¿Cómo se conecta lo que C7 afirma con el

objetivo general de esta investigación? Dicho de otro modo: ¿Facilita el hallazgo anunciado por C7 un dato valioso para decidir la polémica entre cognitivistas y no-cognitivistas *morales*? En mi investigación he preferido ser prudente a la hora de responder a esta cuestión general. He dejado en suspenso hasta qué punto lo afirmado por C7, una tesis sobre el estatuto de nuestra capacidad para ser racionales, podría decidir el debate entre cognitivistas morales y no cognitivistas *morales*. Hay dos razones generales que me han impulsado a adoptar esta línea moderada.

La primera razón es de índole estructural. Como ya he señalado, el ideal que mueve esta investigación gira en torno a la posibilidad de acometer un dominio teórico problemático (*Meta-ética*) a partir de ciertos cuestiones que surgen en un dominio aparentemente menos contencioso (*Meta-racionalidad*). Para sustentar esta reducción metodológica he apelado a un foco de interés común, nuestro concepto normativo de *razón*. Tanto nuestras reglas morales como nuestras reglas de racionalidad, he sostenido, sustentan su normatividad en torno a cierto tipo de *razones* que podrían justificar nuestra conducta en un determinado contexto (NR arriba). A pesar de la solidez que le presupongo a la tesis *racionalista*, seguramente alguien estaría dispuesto a mantener que deberíamos hilar más fino si quisiéramos extraer conclusiones *meta-éticas* fiables a partir de ciertos hallazgos provenientes de lo que aquí he venido denominando *Meta-racionalidad.* Dos vías adicionales de semejanza, defienden, deberían investigarse para motivar la importancia de C7 dentro del debate meta-ético reciente. Por un lado, deberíamos preguntarnos si los hechos normativos *morales* estructuran nuestra deliberación del mismo modo que aquellos hechos normativos ligados a nuestra *racionalidad*. Por otro lado, no estaría de más saber si las prácticas mediante las que atribuimos responsabilidad moral comparten el mismo tipo de regularidades (a nivel personal, social e institucional) que aquellas por las que juzgamos el mérito de los agentes al formar o revisar sus actitudes de acuerdo con ciertas reglas de racionalidad. Si después de investigar detalladamente estas dos dimensiones seguimos encontrando similitudes entre los dos tipos de hechos normativos, lo que C7 señala podría fijar una conclusión válida para la *Meta-ética.* C8 se hace eco de esto

- **(C8)** C7 podría conformar un dato relevante para la *Meta-ética* sólo si puede complementarse con un análisis más detallado de las semejanzas entre nuestra

receptividad moral y el tipo de receptividad que expresamos al responder a una razón derivada de una regla de racionalidad

Como ya adelanté tengo una razón adicional para favorecer una interpretación prudente de C7. Si estoy en lo cierto, el mismo tipo de complejización que motivó el *impasse* a nivel *meta-ético* se puede replicar en el ámbito propio de la *Meta-racionalidad*. Así como la posibilidad del *impasse a nivel meta-ético* está ligada a la disponibilidad de ciertas versiones *sofisticadas* de *expresivismo*

> **(Expresivismo sofisticado)** *Expresivismo* más la tesis de que nuestras opiniones morales se profieren convencionalmente junto a cierto tipo de creencias (*creencias morales*)

en la parte final de mi tesis asumo que el mismo tipo de complejización podría ejemplificar una situación de indefinición dentro del ámbito representado por la *Meta-racionalidad*. Por ponerlo de modo intuitivo: si un expresivista *sofisticado* pudiese demostrar que cumplimos con ciertas reglas básicas de racionalidad mediante ciertas creencias normativas de segundo nivel, seguramente tendríamos que tomar con cierto escepticismo el proyecto metodológico esbozado en este trabajo. Y esto es así porque la epítome de este proyecto, C7, no facilita una ruta directa para resolver el *impasse* en *Meta-ética*. C7 anunciaría, como mucho, la necesidad de profundizar aún más en el enfoque oblicuo que anima esta investigación – al menos si queremos superar esta nueva indefinición. Pero C7 no solucionaría pos sí mismo el impasse en el seno de la *Meta-ética*. Al fin y al cabo, si un análisis expresivista de nuestras capacidades racionales puede acomodar C7, ¿qué nos puede aportar lo que C7 dice para resolver el *impasse* en *Meta-ética*? En suma, si lo que C7 afirma puede ser acomodado, como parece, por aquellos que defienden un expresivismo normativo generalizado, C7 constituye una base insuficiente para reiniciar la *Meta-ética* de modo concluyente.

Una vez que asumimos este pesimismo, no obstante, ¿qué ruta o enfoque debemos favorecer? ¿Debemos recular y cuestionarnos de nuevo la validez de los argumentos motivacionales - o de aquellos que parten de una interpretación sustantiva del concepto de *razón* en sentido normativo? ¿Debemos prestar más atención a aquellos puzzles semánticos que tanto preocuparon a Peter Geach? ¿Debemos atender, en otro orden de cosas, a los datos

que facilitan ciertos enfoques empíricos? ¿Convertirnos, quizás, en filósofos experimentales? A pesar de lo que el ritmo retórico pudiera sugerir, ninguna de las rutas sugeridas por estas preguntas me parece enteramente implausible. Es más, la mayoría resultan tremendamente atractivas a la luz de algunos desarrollos recientes. En cualquier caso - y en la medida en que el tono general de mi tesis está a medio camino entre la *Meta-ética* y la *Teoría de la Racionalidad* - estoy más interesado en otro tipo de alternativa. Tal y como veo el debate, dos rutas alternativas son plausibles para fijar el contenido de C7. Ambas rechazan un enfoque meta-ético estándar (favoreciendo un acercamiento oblicuo a la naturaleza de nuestros juicios morales).

En primer lugar, un posible modo de distinguir entre lo que un cognitivista y un expresivista podrían afirmar sobre nuestra capacidad para ser racional a través del ejercicio de nuestra capacidad para formar creencias normativas de segundo nivel podría apelar a un análisis de nuestro concepto de *creencia*. Tomando el concepto de *creencia* como foco, podríamos comenzar por iluminar su contenido conceptual. Una vez logrado un acuerdo sobre aquellos rasgos básicos que gobernarían el uso de este concepto, podríamos preguntar si las posibles interpretaciones de C7 (la interpretación cognitivista y la interpretación expresivista) pueden acomodar con igual precisión nuestro concepto de *creencia*, tal y cómo esté se define a partir de un proceso de análisis conceptual. Este proceso indagaría, por ejemplo, en las regularidades psicológicas y fenomenológicas que definen nuestro concepto de *creencia*. A partir de ahí nos podríamos preguntar si la totalidad de esos rasgos podrían ser recogidos por alguna de nuestras teorías meta-éticas. Esto, evidentemente, no es una cuestión de todo o nada. Pero podemos convenir que aquella teoría que no sea capaz de acomodar los rasgos más centrales de nuestro concepto de *creencia* tendrá un punto en su contra. Igualmente, aquella teoría que logre acomodar esos rasgos centrales será tremendamente plausible. Plausible a un nivel donde la otra teoría fracasa. La teoría más comprehensiva podría ofrecer, por tanto, una mejor interpretación de C7. Y a partir de esta interpretación, una base fiable desde la cual derivar una conclusión propiamente meta-ética.

El enfoque que acabo de esbozar resultará atractivo para muchos. En la parte final de mi tesis, sim embargo, favorezco una ruta. Las razones que aduzco para justificar mi proceder son bastante directas. En la medida en que la solución esbozada arriba depende de que seamos capaces de delinear una teoría general acerca de nuestras *creencias*, cancelar la indefinición implícita en C7 según esta ruta supondría renunciar a un dominio mínimo de autonomía para

la *Meta-ética*. Por tanto, en la medida en que mi trabajo se ubica decididamente dentro de los límites de la *Meta-ética*, la solución que acabo de esbozar queda sin explorar en estas páginas. Según creo, aquellos conceptos básicos que estructuran el debate meta-ético reciente (*creencia*, *representación*, *verdad*, *aserción*, *coordinación*) deben estudiarse dentro del único fenómeno que unifica cualquier ámbito normativo. Me refiero, por supuesto, a nuestra capacidad para deliberar y al modo en que esta capacidad posibilitaría nuestras respuestas a lo que he venido denominando como *normatividad* . Por tanto, la ruta que defiendo es clara: Si queremos decidir qué teoría explica mejor C7, atendamos a nuestra deliberación en aquellos contextos en los que una regla de racionalidad se aplica. Si logramos delinear, en suma, un argumento por el cual una determinada teoría meta-normativa centrada en el estatuto psicológico de nuestra capacidad racional (M) es capaz de ofrecer una buena explicación de nuestra deliberación en ciertos contextos, entonces M debería asumirse como la mejor interpretación de C7. En mi tesis he abogado por dos argumentos para delimitar el contenido de C7.

De un lado, podríamos preguntar por las implicaciones que acarrearía la aceptación de una determinada teoría meta-normativa (M) para el tipo de *estabilidad* que presuponemos al cumplir deliberativamente con lo demandado por nuestras reglas de racionalidad. Reparemos, por ejemplo, en aquellos procesos psicológicos que deberíamos ser capaces de instanciar si cumpliéramos con una determinada regla de racionalidad en sentido amplio. Estos procesos requerirían un cierto grado de *estabilidad* por parte de aquellos estados psicológicos que sustentan nuestro razonamiento. Supongamos que ese tipo de *estabilidad* podría explicarse mejor si presupusiéramos una lectura cognitivista de M. Por tanto, si pudiera defenderse que la *estabilidad* necesaria para soportar nuestro razonamiento requiere una lectura cognitivista de nuestras creencias de segundo nivel, entonces podríamos colegir que C7 sustenta una tesis cognitivista a nivel meta-normativo.

Este argumento no es concluyente. Un expresivista sofisticado podría aceptar que nuestras procesos de revisión de actitudes requieren cierta *estabilidad.* Aceptar esto no le exigiría explicar esa *estabilidad* apelando a una creencia centrada en una determinada relación meta-proposicional ('ser demandado', 'ser requerido'). Para este expresivista sofisticado, otro criterio de *estabilidad* sería posible. ¿Cuál? En la medida en que ciertos estados no-cognitivos – *decisiones*, *planes* o *intenciones* – son objeto de desacuerdo, nuestro expresivista defiende que esos estados psicológicos no-cognitivos pueden replicar todas las regularidades

deliberativas que una interpretación *cognitivista* dice acomodar mejor. Una lectura cognitivista de C7 podría jugar la carta de la *estabilidad*, por tanto, sólo si explicase por qué su sentido de *estabilidad* es superior al sentido sugerido por el expresivista. En cualquier casos si el mismo conjunto de regularidades deliberativas pueden acomodarse a partir de dos interpretaciones incompatibles de *estabilidad*, resulta complicado argüir que tenemos un argumento encaminado a precisar C7 a partir de un análisis del tipo de *estabilidad* que presuponemos en nuestra deliberación

Pero en cualquier caso lo anterior no representa la ruta que favorezco en la parte final de mi tesis. Allí apunto que un modo plausible especificar el tipo de creencia mencionada en C7 debe centrarse en nuestra deliberación tal y cómo nosotros la percibimos. Es decir, podemos precisar qué tipo de creencia ejemplificamos al deliberar en sentido consciente si nos centramos en la fenomenología de la deliberación. En este dominio, sugiero, la explicación expresivista resultar claramente contra intuitiva. El tipo de creencia mencionada en C7 no debería interpretarse en clave expresivista si queremos respetar la *necesidad* que el agente experimenta al deliberar.

Cuando nos centramos en nuestra perspectiva deliberativa una cosa parece clara: como deliberadores no experimentar la necesidad impuesta por nuestras reglas de racionalidad como algo dependiente de nuestras elecciones. Si creemos que debemos formar cierta intención, por ejemplo, la necesidad deóntica que experimentamos no se percibe desde nuestra perspectiva como algo dependiente de una determinada política práctica, o de un compromiso con un determinado sistema de normas. Más bien percibimos ese tipo de *necesidad* como algo que tiene que ver con la verdad de ciertos hechos. Percibimos, por decirlo de modo intuitivo, que sólo si creemos que p y que si p entonces q es verdad que debemos creer que q. Pero si esto es así – si la creencia que tiene como contenido un principio de racionalidad WS dirige nuestra indagación de primer nivel atendiendo a la verdad o falsedad de ciertos hechos psicológicos – entonces ¿por qué no afirmar que creemos algo cuando deliberamos desde nuestras creencias de segundo nivel? Si en ningún caso la necesidad que rastrea nuestra creencia de segundo nivel se percibe como algo relativo a una determinada política cognitiva, como los expresivistas señalan, ¿por qué no desechar simplemente la explicación ofrecida por el expresivismo sobre nuestra capacidad para deliberar? Si no experimentamos como algo dependiente de nuestra elección el hecho de que si es verdad que creemos que p y que si p

entonces q, entonces debemos creer que q, ¿por qué no abrazar el cognitivismo para explicar el tipo de creencia a la que C7 refiere?

En mi trabajo defiendo que este recurso a la fenomenología de la deliberación tampoco establece de modo concluyente una interpretación cognitivista de C7. Esto se debe a que los expresivistas pueden acomodar también este sentimiento o rasgo fenomenológico. O al menos lo pueden acomodar en sentido externo, es decir, pueden explicar la función de ese sentimiento en un marco expresivista general. Según ellos, lo que en algunos contextos deliberativos percibimos como algo demandado por un hecho externo, un hecho independientes de cualquier decisión o política contingente, no es más que una faceta más de nuestros compromisos con *normas de segundo nivel*. Las normas de segundo nivel prescriben el mismo conjunto de disposiciones y regularidades que normalmente asociamos con nuestra fenomenología deliberativa, es decir, con la imagen manifiesta que se nos presenta como obvia al deliberar. Esas disposiciones y regularidades son explicadas por los expresivistas sin presuponer ningún dominio de hechos y ninguna capacidad de tipo perceptivo. Simplemente ejemplificamos ese conjunto de disposiciones, regularidades conductuales y reacciones debido a nuestro compromiso con un sistema de normas epistémicas de segundo nivel – y estas normas no pretenden describir ningún dominio de hechos sino favorecer la coordinación intrapersonal y extrapersonal o social. Incluso nuestra percepción fenomenológica sería reforzada por esas normas y por las respuestas de otros agentes ante casos de incumplimiento. En ningún caso, por tanto, responderíamos literalmente a un hecho normativo al comportarnos racionalmente, es decir, al responder a una razón epistémica. Y en ningún caso deberíamos presuponer una creencia para explicar nuestra fenomenología al deliberar. Todo puede acomodarse, señala el expresivista, dentro de un mismo molde semántico, metafísico y epistemológico que no presupone una noción de objetividad

La última conclusión que se extrae de mi investigación, por tanto, afirma que

- **(C9)** C7 podría conformar un dato relevante para la *Meta-ética* sólo si estableciésemos qué tipo de actitud (*creencia* o *quasi-creencia*) estructura nuestras respuestas ante ciertas razones epistémicas. Esto equivale (1) a indagar de modo general en nuestro concepto de *creencia* y *representación* o **(2)** a preguntar por las

aquellas condiciones psicológicas que sustentarían las regularidades y la fenomenología inherente a nuestra perspectiva deliberativa.

REFERENCES

- Altham, J. E. 1986. "The Legacy of Emotivism", in McDonald, G. Wright, C. (eds). *Facts, Science, and Morality; Essays on A. J. Ayer's Language, Truth, and Logic*, Basil Blackwell, Oxford

- Anscombe, G. E. M., 2000/1963, *Intention*, Harvard, Harvard University Press.

- Arpaly, N. 2003. *Unprincipled Virtue*, Oxford, Oxford University Press.

- Audi, R. 2006. *Practical Reasoning and Ethical Decision*, London, Routledge

- Ayer, A. J. 2001/1936. *Language, Truth, and Logic*, London, Penguin

- Bennett, J. 1995. *The Act Itself*, Oxford, Oxford University Press

- Bermudez, J. L., Millar, A. (eds), 2002, *Reasons and Nature. Essays in the Theory of Rationality,* Oxford University Press, Oxford.

- Bittner, R., 2001, *Doings Things for Reasons,* Oxford University Press, Oxford.

- Blackburn, S. 2006. "Anti-Realist Expressivism and Quasi-Realism", in Copp, D. (ed). 2006

- ------, 2006a. "Must We Weep for Sentimentalism?", in Dreier, J. (ed). 2006. pp. 144-161.

- ------, 1998. *Ruling Passions*, Oxford, Oxford University Press

- ------, 1993. *Essays in Quasi-Realism*, Oxford, Oxford University Press

- ------, 1993/1990. "Just Causes", originally in Philosophical Studies, Vol. 61, pp. 3-42

- ------, 1992. "Wise Feelings, Apt Reading". Ethics, 102, January, pp. 342-356

- ------, 1988/1993. "How to Be an Ethical Antirealist?", originally in Midwest Studies, Vol. 59

- ------, 1993/1985. "Errors and the Phenomenology of Value", originally in Honderich, T. (ed). *Ethics and Objectivity*, London, Routledge & Kegan Paul.

- ------, 1984. *Spreading the Word*, Oxford, Oxford University Press.

References

- Beall, J. Restall, G. 2005."Logical Consequence" in Eward N. Zalta (ed), *The Stanford Encyclopedia of Philosophy*, Winter 2003 Edition, http://plato.stanford.edu/entries/logical-consequence/.

- Boyd, R. 1988. "How to Be a Moral Realist" in Sayre-McCord, G. (ed). 1988.

- Brandt, R. 1979. *A Theory of Good and the Right,* Cambridge, Cambridge University Press

- Bratman, M. 2007. *Structures of Agency*, Oxford, Oxford University Press

- ------, 2007/2004. Shared Valuing and Framework for Practical Reasoning", originally in Wallace, R. J. Pettit, P. Scheffler, S. Smith, M. (eds). 2004. pp. 1-28

- ------, 2007/2000. "Reflection, Planning, and Temporally Extended Agency", originally in The Philosophical Review, 109, pp. 35-61

- ------, 2007/2000b. "Valuing and the Will", originally in Philosophical Perspectives 14, pp. 249-265

- Bratman, M. 1999. *Faces of Intention*, Cambridge, Cambridge University Press

- ------, 1999/1992. "Practical Reasoning and Acceptance in Context", originally in Mind, 101, pp. 1-14

- ------, 1987. *Intention, Plans, and Practical Reasons,* California, CSLI Publications

- Brennan, G. Goodin, R. Jackson, F. Smith, M. (eds). 2007. *Common Minds; Themes from the Philosophy of Philip Pettit*, Oxford, Oxford University Press.

- Bricke, J. 1996, *Mind and Morality. An Examination on Hume's Moral Psychology*, Oxford University Press, New York.

- Brink, D. O. 2007. "Review of Allan Gibbard's *Thinking How to Live*", Philosophical Review, Vol. 116, n° 2

- ------, 1989. *Moral Realism and the Foundations of Ethics,* Cambridge, Cambridge University Press.

- Broome, J. (forthcoming _1). "The Unity of Reasoning"

- ------. (forthcoming _2). "Is Rationality Normative?"

- ------, 2007. "Requirements", in Rønnow-Rasmussen, T. Petersson, B. Josefsson, J. and Egonsson, D. *Homage à Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*, http://www.fil.lu.se/HommageaWlodek/site/abstra.htm.

- ------, 2007a. "Wide or Narrow Scope?", <u>Mind</u>, 107, pp. 359-370

- ------. 2006. "Reasoning with preferences?", in Olsaretti, S. (ed). 2006.

- ------. 2005. "Does Rationality Give us Reasons?", <u>Philosophical Issues</u>, 15, 321-337

- ------. 2004. "Reasons" in Wallace, R. J. Pettit, P. Scheffler, S. Smith, M. (eds). 2004

- ------, 2002, "Practical Reasoning", en Bermudez, J. L. Millar, A.(ed), 2002, pp. 85-113.

- ------, 2001, "Normative Practical Reasoning", *Proceedings of the Aristotelian Society*, Supp. Vol. Nº 75: 175-193.

- ------. 2001a. "Normative Requirements" en Dancy, J. (ed). *Normativity*, Blackwell

- ------, 1997. "Reasons and Motivation", <u>Proceedings of the Aristotelian Society</u>, Vol. 71

- Burge, T. 1998. "Reasons and the First-Person", in Wright, C. Smith, B. Macdonald, C. (eds). *Knowing Our Own Minds*, Oxford, Oxford University Press

- Christiansen, D. 2004. *Putting Logic in its Place. Formal Constraints on Rational Belief*, Oxford, Oxford University Press

- Copp, D. 2007. *Morality in a Natural World. Selected Essays in Meta-Ethics*, Cambridge, Cambridge University Press

- ------, 2007/2003. "Why Naturalism?", originally in <u>Ethical Theory and Moral Practice</u>, 6, pp. 179-200

- ------. (ed), 2006, *Oxford Handbook on Ethical Theory*, Oxford, Oxford University Press.

- Chrisman, M. (forthcoming). "Expressivism, Inferentialism, and Saving the Debate"

- Crisp, R. 2006. *Reasons and the Good*, Oxford, Oxford University Press

- Cullity, G. Gaut, B. (eds). 1997. *Ethics and Practical Reasons*, Oxford University Press, Oxford

References

- Cuneo, T. 2007. *The Normative Web. An Argument for Moral Realisms*, Oxford, Oxford University Press

- ------, 2006. "Moral Facts as Configuring Causes", Pacific Philosophical Quarterly, 87, pp. 141-162

- ------, 2006a. "Saying what we Mean. An Argument against Expressivism, in Shafer-Landau, R. (ed). 2006. pp. 35-73

- Damasio, A. 1994. *Descartes' Error*, London, Vintage.

- Dancy, J. (forthcoming). "Reasons and Rationality"

- ------, 2006. "Non- naturalism", in Copp, D. (ed). 2006. pp. 122-146.

- ------, 2004, *Ethics without Principles*, Oxford, Oxford University Press.

- ------, (ed), 2001, *Normativity*, Blackwell, London.

- ------, 2000, *Practical Reality*, Oxford University Press, Oxford.

- ------, 1994, "Why there is Really no such Thing as a Theory of Motivation?", *Proceedings of the Aristotelian Society*, Vol. 95, pp. 1-18.

- ------, 1993, *Moral Reasons*, Blackwell, London.

- Daniels, N. "Wide Reflective Equilibrium and Theory Acceptance in Ethics", Journal of Philosophy, 76, pp. 256-282

- Darwall, S. 2006. "How Should Ethics Relate to (the Rest of) Philosophy? Moore's Legacy", in Horgan, T. Timmons, M. 2006

- ------, 1983, *Impartial Reason,* Cornell University Press, Ithaca.

- Darwall, S. Gibbard, A. Railton, P. 1992. "Toward Fin de siècle Ethics: Some Trends", Philosophical Review, 101, pp. 115-189

- Davidson, D., 1963, "Action, Reasons, and Causes" en Mele, A., 1997, *The Philosophy of Action*, Oxford University Press, Oxford, pp. 27-42

- De Paul, M. "Intuitions in Moral Inquiry", in Copp, D. (ed). 2006. pp. 595-624

- Divers, J. Miller, A. 1994. "Why Expressivists about Value should Not Love Minimalism about Truth?", <u>Analysis</u>, 54, pp. 12-19

- Dorr, C. 2002. "Non-cognitivism and Wishful Thinking", <u>Noûs</u>, 36, Nº 1, pp. 97-103

- Dreier, J. (ed). 2006. *Contemporary Debates in Moral Theory*, Oxford, Blackwell.

- ------, 2006a. "Was Moore a Moorean?", in Horgan, T. Timmons, M. (eds). 2006. pp. 191-209

- ------, 2004. "Meta-ethics and the Problem of Creeping Minimalism", <u>Philosophical Perspectives</u>, 18

- ------, 2006/2000. "Dispositions and Fetishes: Externalist Models of Moral Motivation", in Fisher, A. Kirchin, S. (eds). 2006. pp. 547-567, originally in <u>Philosophy and Phenomenological Research</u>, 61, 619-638

- ------, 1997. "Humean Doubts about Practical Reasons", in Cullity, G. Gaut, B. (eds). 1997

- D'Arms, J. Jacobson, D. 2006. "Sensibility Theory and Projectivism", in Copp, D. (ed). 2006. pp. 186-218

- ------. 2000. "Sentiment and Value", <u>Ethics</u>, 110, pp. 722-748.

- Enoch, D. "An Outline of An Argument for Robust Metanormative Realism", in Shafer-Landau, R. (ed). 2007. pp. 21-51

- Falk, W. D. 1948. "Ought and Motivation", *Proceedings of the Aristotelian Society*, 48, (1947-1948), pp. 492-510

- Finlay, S. 2007. "Responding to Normativity", in Shafer-Landau, R. (ed). 2007

- Fischer, J. M. Ravizza, M. 1998. *Responsibility and Control,* Cambridge, Cambridge University Press

- Fisher, A. Kirchin, S. (eds). 2006. *Arguing about Metaethics*, London, Routledge

- Foot, P. 2002, *Virtues and Vices,* Oxford, Oxford University Press.

- ------, 2001. *Natural Goodness*, Oxford, Oxford University Press

- ------, (ed). 1974. *Theories of Ethics*, Oxford, Oxford University Press

References

- ------, 2002/1972, "Morality as a System of Hypothetical Imperatives", originally in <u>The Philosophical Review</u>, Vol. 81, nº 3.

- ------, 2002/1961. "Goodness and Choice", originally in <u>The Aristotelian Society Supplementary Volume</u>, 1961

- ------, 2002/1958-1959. "Moral Beliefs", originally in <u>Proceedings of the Aristotelian Society</u>, Vol. 59

- ------, .2002/1958. "Moral Arguments", originally in <u>Mind</u>, Vol. 67.

- Frankena, W. 1951. "Moral Philosophy at Mid-Century", <u>The Philosophical Review</u>, Vol. 60, Nº 1, pp. 44-55.

- ------, 1958. "Obligation and Motivation in Recent Moral Philosophy" in Melden, A. I. (ed). *Essays on Moral Philosophy*, Washington, University of Washington Press.

- ------, 1974/1939. "The Naturalistic Fallacy", originally in <u>Mind</u>, Vol. 48, pp. 464-467, reprinted in Foot, P. (ed). 1974. pp. 50-63

- Frankfurt, H. 2006. *Taking Ourselves Seriously & Getting it Right. The Tanner Lectures*, Stanford, Stanford University Press

- ------, 2005. *The Reasons of Love*, Princeton, Princeton University Press

- ------, 1971. Freedom of the Will and the Concept of Person", <u>Journal of Philosophy</u>, 68, pp. 5-20

- Frey, R. G. 2002. Act-Utilitarianism" in LaFollette, H. (ed). 2002. pp. 165-183

- Gaitán, A. 2005. "Pendiendo de una dudosa cuerda; El acceso normativo en Practical Reality", <u>Crítica</u>, Vol. 37, Nº 109, pp. 65-97

- Geach, P. 1974/1956. "Good and Evil", originally in <u>Analysis</u>, vol. 17. pp. 33-42, reprinted in Foot, P. (ed). 1974. pp. 64-73

- Gauthier, D. 1986. *Morals by Agreement*, Oxford, Oxford University Press

- Gewirth, A. 1960. "Meta-ethics and Normative Ethics", <u>Mind</u>, LXIX, April, 187-205

- Gibbard, A. 2006. "Précis of *Thinking How to Live*", <u>Philosophy and Phenomenological Research</u>, Vol. LXXII, nº 3, May

- ------, 2003. *Thinking How to Live*, Cambridge, Harvard University Press

- ------. 2007/2002. "The Reasons of a Living Being", originally in <u>Proceedings of the American Philosophical Association</u>, 62, reprinted in Shafer-Landau, R. Cuneo, T. (eds). 2007. pp. 71-78

- ------, 1990. *Wise Choices, Apt Feelings*, Cambridge, Harvard University Press

- Greene, J. Haidt, J. 2002. How (and where) does moral judgment work?", <u>Trends in Cognitive Sciences</u>, 6, pp. 517-523

- Greenspan, P. 2007. "Practical Reasons and Moral 'Ought'", in Shafer-Landau, R. (ed). 2007. pp. 172-194

- ------, 1975. "Conditional Oughts and Hypothetical Imperatives", <u>Journal of Philosophy</u>, 72, pp. 259-276

- Hare, R. M. 1952. *The Language of Morals*, Oxford, Oxford University Press

- ------, 1963. *Freedom and Reason*, Oxford, Oxford University Press

- ------, 1981. *Moral Thinking*, Oxford, Oxford University Press

- Harman, G. 1998/1977. "Ethics and Observation", originally in *The Nature of Morality*, Chapter 1, reprinted in Rachels, J. (ed). 1998

- ------, 1997/1976, "Practical Reasoning", originally in <u>Revue of Metaphysics</u>, 79, 431-463, reprinted in Mele, A. (ed). 1997.

- ------, 1986, *Change in View*, MIT Press, Cambridge.

- Harman, G. Thompson, J. J. 1996. *Moral Relativism and Moral Objectivity*, Oxford, Blackwell

- Heuer, U. 2004. "Raz on Values and Reasons", in Wallace, R. J. Pettit, P. Scheffler, S. Smith, M. (eds). 2004. pp. 129-153.

References

- Hussain, N. Shah, N. 2006. "Misunderstanding Meta-ethics; Korsgaard's Rejection of Realism", in Shafer-Landau, R. (ed). 2006. pp, 265-295

- Hieronymi, P. 2005. "The Wrong Kind of Reasons", <u>The Journal of Philosophy</u>, Vol. 102, nº 9, pp. 437-457

- ------. 2006. "Controlling Attitudes", <u>Pacific Philosophical Quarterly</u>, 87

- Hooker, B. 1987. Williams's argument against External Reasons", <u>Analysis</u>, 47, 42-44

- Hooker, B. Streumer, B. 2004. "Substantive and Procedural Practical Rationality", in Mele, A. Rawling, P. (eds). 2004. pp. 57-75

- Holton, R. 2006. "The Act of Choice", <u>Philosophers' Imprint</u>, Vol. 6. Nº 3. September

- Horgan, T. Timmons, M. 2007/2000. "Nondescriptivist Cognitivism; Framework for a New Metaethic", originally in <u>Philosophical Papers</u>, 29, pp. 121-153, reprinted in Shafer-Landau, R. Cuneo, T. 2007.

- ------, (eds). 2006. *Metaethics after Moore*, Oxford, Oxford University Press

- ------. 2006a. "Morality without Moral Facts". In Dreier, J. (ed). 2006

- ------, 2006b. "Cognitivist Expressivism", in Horgan , T. Timmons, M. (eds). pp. 255-298

- Hornsby, J.,1997. *Simple Mindedness; A Defence of Naïve Naturalism in the Philosophy of Mind*, Harvard, Harvard University Press.

- Horwich, P. 2005. "The Frege-Geach Point", <u>Philosophical Issues</u>, 15, pp. 78-93

- ------, 1993. "Gibbard's Theory of Norms", <u>Philosophy and Public Affairs</u>, 22/1, 67-78

- ------, 1990. *Truth*, Oxford, Oxford University Press

- Humberstone, I. L., 1992, "Direction of Fit", <u>Mind</u>, 101, 59-83

- Hume, D. 1978. *A Treatise of Human Nature*, Oxford, Oxford University Press.

- Hurka, T. 2003. "Moore in the Middle", <u>Ethics</u>, 113, pp. 599-628

- Jackson, F. 2003. "Cognitivism, A Priori Deduction, and Moore", <u>Ethics</u>, 113, nº 3

- ------, 2000. "Non-cognitivism, Normativity, Belief", in Dancy, J. (ed). 2000. pp. 100-116

- ------, 2000. "What is Expressivism?", <u>Philosophical Books</u>, 42, (1), pp. 10-17

- ------. 1998. *From Metaphysics to Ethics; A Defence of Conceptual Analysis*, Oxford, Oxford University Press

- Jackson, F. Oppy, G. Smith, M. 1994. "Minimalism and truth-aptness", Mind, 103, pp. 287-302

- Jackson, F. Pettit, P. 1998. "A Problem for Expressivism", <u>Analysis</u>, 58, pp. 239-51

- ------, 1995. "Moral Functionalism and Moral Motivation", <u>Philosophical Quarterly</u>, 45, pp. 20-40

- ------, 1990. "Program Explanation; A General Perspective", <u>Analysis</u>, 50, pp. 107-117

- Joyce, R. 2006. *The Evolution of Morality*, Cambridge, MIT Press

- ------, 2002. *The Myth of Morality,* Cambridge, Cambridge University Press

- Kagan, S. 1994. *The Limits of Morality*, Oxford, Clarendon Press

- Kalderon, M. E. 2005. *Moral Fictionalism*, Oxford, Oxford University Press

- Kamm, F. M. 2006. *Intricate Ethics*, Oxford, Oxford University Press.

- Kavka, G. 1983. "The Toxin Puzzle", <u>Analysis</u>, 43, 33-36

- Kelly, T. 2006. "The Epistemic Significance of Disagreement", in Hawthorne, J. Gendler, T. T. (eds). *Oxford Studies in Epistemology, Vol.* 1, Oxford, Oxford University Press

- Kolodny, N., 2007. "States or Processes Requirements?", <u>Mind</u>, 116 (462), pp. 371-385

- -------, 2005. "Why be Rational?", <u>Mind</u>, 114, pp. 509-63

- Korsgaard, C. M. 1986. "Skepticism about Practical Reason", <u>Journal of Philosophy</u>, 83, pp. 5-25

- ------, 1997, "The Normativity of Practical Reason", in Cullity, G. Gaut, B. (ed), 1997.

- ------, 1998. *The Sources of Normativity*, Cambridge, Cambridge University Press

References

- Kripke, S. 1972. *Naming and Necessity*, Cambridge, Harvard University Press

- LaFollette, H. (ed). 2002. *The Blackwell Guide of Ethical Theory*, Oxford, Blackwell.

- Lance, M. Little, M. 2006. "Particularism and Anti-theory", in Copp, D. (ed). pp. 567-595

- ------, 2006a. "Defending Moral Particularism" in Dreier, J. (ed). 2006. pp. 305-323

- Leiter, B. (ed). 2004. *The Future for Philosophy*, Oxford, Oxford University Press

- ------, 2001. (ed). *Objectivity in Law and Morals*, Cambridge, Cambridge University Press

- ------, 2001a. "Moral Facts and Best Explanations", <u>Social Philosophy & Policy</u>, Vol. 18. Summer, pp. 79-101

- Lenman, J. 2003. "Non-cognitivism and Wishfulness", <u>Ethical Theory and Moral Practice</u>, Nº 6, pp. 265-274

- Little, M. O. 1997. "Virtue as Knowledge; Objections from the Philosophy of Mind", <u>Noûs</u>, 31, pp. 59-77

- McDonald, C. McDonald, G. (eds). 2006. *McDowell and his Critics*, Oxford, Blackwell.

- McDowell, J. 1998. *Mind, Value, and Reality*, Cambridge, Harvard University Press

- ------, 1998a, "Might Be There External Reasons?", in McDowell, J. 1998. pp. 95-112

- 1998/1979. "Virtue and Reason", pp. 50-77, originally in <u>The Monist</u>, 62, pp. 331-350

- 1998/1978. "Are Moral Requirements Hypothetical Imperatives?", originally in Proceedings of the Aristotelian Society, Vol. 52, pp. 13-29

- McKeever, S. Ridge, M. 2006. *Principled Ethics. Generalism as a Regulative Ideal*, Oxford, Oxford University Press.

- McNaughton, D., 1988, *Moral Vision*, Oxford, Basil Blackwell

- McNaughton, D., Rawling, P., 2004, "Duty, Rationality and Practical Reason", in Mele, A. R., Rawling, P. (eds.), 2004. pp. 110-131.

- ------, 2003. "Naturalism and Normativity", <u>Proceedings Aristotelian Society Supplementary Volume</u>, 77, 1, pp. 23-45

- Mele, A. R., 2005, "Action", in Jackson, F., Smith, M., 2005.

- ------, 2003, *Motivation and Agency,* Oxford, Oxford University Press

- ------, 1997. (ed). *Philosophy of Action*, Oxford, Oxford University Press.

- Mele, A. Rawling, P. 2004. (eds). *Oxford Handbook of Rationality*, Oxford, Oxford University Press

- Millar, A., 2002, "Reasons for Action", en Bermudez, J. L. Millar, A. 2002, pp. 113-135.

- Miller, A. 2003. *An introduction to Contemporary Meta-ethics*, London, Polity Press

- Moore, G. E. 1942. "A Reply to my Critics", in Schilpp, P. A. 1942. The Library of Living Philosophers. Vol. VI, Evanston, Northwestern University Press,

- ------, 2005/1912. *Ethics*, Oxford, Oxford University Press

- ------, 1993/1903. *Principia Ethica*, Cambridge, Cambridge University Press.

- Moran, R., 2003, *Authority and Estrangement*, Princeton, Princeton University Press.

- Nagel, T., 1986, *The View from Nowhere*, Clarendon Press, Oxford.

- ------, 1970, *The Possibility of Altruism*, Princeton University Press, Princeton.

- Nahmias, E. et al. 2004. "The Phenomenology of Free Will", Journal of Consciousness Studies, 11, pp. 162-179

- Nichols, S. 2004. *Sentimental Rules. On the Natural Foundations of Moral Judgments*, Oxford, Oxford University Press

- Nuccetelli, S. Seay, G. (eds). 2007. *Themes from G. E. Moore*, Oxford, Oxford University Press.

- Olsaretti, S. 2006. (ed). *Preferences and Well-Being*, Cambridge, Cambridge University Press

- Parfit, D., 2006, "Normativity" in Shafer-Landau, R. (ed), 2006, pp. 325-381

- ------, 2001, "Rationality and Reasons", in Egonsson, D., Josefsson, J., Petersson, B., Ronnow-Rasmussen, T. (eds.), *Exploring Practical Philosophy: From Action to Values*, Ashgate, pp. 17-39.

References

- ------, 1997, "Reasons and Motivation", <u>Proceedings of the Aristotelian Society</u>, Vol 71, pp. 99-130.

- ------, 1984, *Reasons and Persons*, Oxford University Press, Oxford.

- Pettit, P. 2007. "Joining the Dots" in Brennan, G. Goodin, R. Jackson, F. Smith, M. (eds). 2007. pp. 215-344

- ------, 2007a. "Rationality, Reasoning, and Group Agency", <u>dialectica</u>, Vol. 61, Nº 4, pp. 495-519

- ------, 2006. "On Thinking How to Live: A Cognitivist View", <u>Mind</u>, 115, pp. 1083-1106

- ------, 2004. "Existentialism, Quietism, and the Role of Philosophy", in Leiter, B. (ed). 2004. pp. 304-327

- ------, 2002. *Rules, Reasons, and Norms*, Oxford, Oxford University Press

- ------, 2002a. "Three Aspects of Rational Explanation", in Pettit, P. 2002. pp. 177-192.

- ------. 2001. "Embracing Objectivity in Ethics" in Leiter, B. (ed). 2001.

- Pettit, P. Smith, M. 2006. "Internal Reasons", in McDonald, C. McDonald, G. (eds). 2006.

- ------, 1996. Freedom in Belief and Desire", <u>Journal of Philosophy</u>, 93, pp. 429-449

- Piller, C. 2006. "Content-Related and Attitude-Related Reasons for Preferences" in Olsaretti, S. (ed). 2006. p. 155-183

- ------, 2001. "Normative Practical Reasoning", <u>Proceedings of the Aristotelian Society. Supp. Vol</u>. 75, pp. 195-216

- Prinz, J. 2007. *The Emotional Construction of Morals*, Oxford, Oxford University Press

- Putnam, H. 1975. "The Meaning of Meaning" , in Putnam, H. 1975. *Mind, Language, and Reality*, Cambridge, Cambridge University Press

- Quinn, W. 1993. "Putting Rationality in its Place" in Quinn, W. 1993. *Morality and Action*, Cambridge, Cambridge University Press

- Rabinowicz, W. Ronnow-Rasmussen, T. 2004. "The Strike of the Demon: On Fitting Pro-Attitudes and Value", <u>Ethics</u>, 114, pp. 391-423

- Rachels, J. (ed). 1998. *Ethical Theory. The Question of Objectivity*, Oxford, Oxford University Press

- Railton, P. 2006. "Humean Rationality and Practical Reason", in Copp, D. (ed), 2005, pp. 265-281.

- ------, 2006a. "Normative Guidance" in Shafer-Landau, R. (ed). 2006

- ------, 2006b. "Moral Factualism", in Dreier, J. (ed). 2006

- ------, 2004. *Facts, Values, and Norms*, Cambridge, Cambridge University Press

- ------, 2004a. "How to Engage Reasons: The Problem of Regress", in Wallace, R. J., Pettit, P., Scheffler, S., Smith, M. (eds). 2004. *Reasons and Values: Themes from the Moral Philosophy of Joseph Raz*, Oxford, Oxford University Press,

- ------, 2004/1997. "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action", originally in Cullity, G. Gaut, B. (eds). 1997

- ------, 1996. "Moral Realism; Prospects and Problems", in Sinnott-Armstrong, W. (ed). 1986. pp. 49-82

- ------, 1990. "Naturalism and Prescriptivity" in Frankel, E, Miller, F. Paul, J. (eds). 1990

- ------, 2004/1986, "Moral Realism", originally in <u>Philosophical Review</u>, XCV, Nº 2, pp. 163-207.

- ------, 2004/1986a. "Facts and Values", originally in <u>Philosophical Topics</u>, XIV, nº 2.

- Rawls, J. 1971. *A Theory of Justice*, Cambridge, Harvard University Press

- ------, 1951. "Outline of a Decision Procedure for Ethics", Philosophical Review, 60, pp. 184-186

- Raz, J. 2005. "The Myth of Instrumental Rationality", <u>Journal of Ethics and Social Philosophy,</u> Vol. 1, Issue 1.

- ------. 2001. *Engaging Reasons*, Oxford, Oxford University Press

References

- ------, 2001a. "Agency, Reasons, and the Good", in Raz, J. 2001, pp. 22-45.

- ------, 2001b. "Explaining Normativity: On Rationality and the Justification of Reasons", en Dancy, J. (ed), 2001, pp. 34-59.

- ------, 1999. *Practical Reasons and Norms*, Oxford, Oxford University Press.

- ------, 1978. "Introduction", in *Practical Reasoning*", Oxford, Oxford University Press

- Ridge, M. 2007. "Ecumenical Expressivism: The Best of Both Worlds?", in Shafer-Landau, R. (ed). 2007. pp. 51-77

- ------, 2007a. "Epistemology for Ecumenical Expressivists", <u>Proceedings of the Aristotelian Society. Supplementary Volume</u>, LXXXI, pp. 83-108

- Rosen, G. 1998. "Blackburn's *Essays on Quasi-Realism*", <u>Nous</u>, 32:3, pp. 386-405

- Sayre-McCord, G. 2006. "Moral Realism" in Copp, D. (Ed). 2006. pp. 39-63

- ------ (ed). 1988. *Essays on Moral Realism*, Ithaca, Cornell University Press

- Scanlon, T. 2007. "Structural Irrationality", in Brennan, G. Goodin, R. Jackson, F. Smith, M. (eds). 2007. pp. 84-104

- ------, 2007a. "Wrongness and Reasons: A Re-examination", in Shafer-Landau, R. (ed). 2007a

- ------, 2006. "Reasons and Decisions", <u>Philosophy and Phenomenological Research</u>, Vol. LXXII, nº 3, May

- ------, 2004. "Reasons; A Puzzling Duality?", in Wallace, R. J. Pettit, P. Scheffler, S. Smith, M. 2004.

- ------, 1998. *What we Owe to Each Other*, Harvard University Press, Cambridge.

- Scheffler, S. 1998/1985. "Agent-Centred Restrictions, Rationality, and the Virtues", originally in <u>Mind</u>, 94, pp. 409-419, reprinted in Scheffler, S. (ed). 1998. *Consequentialism and its Critic*s, Oxford, Oxford University Press

- Schueler, G. F., 2003, *Reasons and Purposes. Human Rationality and the Teleological Explanation of Action,* Oxford University Press, New York.

- ------, 1991, "Pro-Attitudes and Direction of Fit", <u>Mind</u>, 100, 277-281

- Setiya, K. 2007. *Reasons Without Rationalism*, Princeton, Princeton University Press

- Shafer-Landau, R. Cuneo, T. 2007. (eds). *Foundations of Ethics*, Oxford, Blackwell

- Shafer-Landau, R. (ed). 2007. *Oxford Studies in Metaethics, Vol. II*, Oxford, Oxford University Press

- ------ (ed), 2006. *Oxford Studies in Metaethics*, *Vol. I*, Oxford, Oxford University Press.

- ------, 2004, *Moral Realism. A Defence,* Oxford, Oxford University Press.

- Shah, N., 2006, "An New Argument for Evidentialism", <u>The Philosophical Quarterly</u>, Vol. 56. 125, pp. 481-498

- Shoemaker, S. 1996. *The First Person Perspective and Other Essays*, Cambridge, Cambridge University Press

- Sinclair, N. (forthcoming). "Presumptive Arguments for Moral Realism"

- ------, 2007. "Propositional Clothing and Belief", <u>The Philosophical Quarterly</u>, Vol. 57. Nº. 228, pp. 342-362

- ------, 2006. "The Moral Belief Problem", <u>Ratio</u>, XIX, Nº 2, June, pp. 249-260

- Sinclair, N. Blackburn, S. 2007. "Comments on Gibbard's Thinking How to Live", <u>Philosophy and Phenomenological Research</u>, Vol. LXXII, May, pp. 699-706

- ------, 2003, "How Truth Governs Belief", <u>Philosophical Review</u>, 112, 447-482

- Sinnott-Armstrong, W. (ed). 1996. *Moral Knowledge?,* Oxford, Oxford University Press

- Skorupski, J. (forthcoming). "What is Normativity?"

- ------, 2001. "Irrealist Cognitivism" in Dancy, J. (ed). 2001. pp. 116-141

- Smith, M., 2005. "Meta-Ethics" in Jackson, F., Smith, M., 2005. 3-31

- ------, 2004, *Ethics and the A Priori. Selected Essays on Moral Psychology and Meta-Ethics*, Cambridge, Cambridge University Press.

References

- ------, 2004a, "Instrumental Desires, Instrumental Rationality", <u>Supplement to the Proceedings of the Aristotelian Society</u>, 78, pp. 93-109.

- ------, 2004b. "Humean Rationality", in Mele, A. Rawling, P. (eds). 2004. pp. 75-92

- ------ 2004/2003, "Humeanism, Psychologism, and the Normative Story", originally in <u>Philosophy and Phenomenological Research</u>, Vol. XVII, nº 2, pp. 460-467.

- ------, 2004/2003a, "Rational Capacities", originally in Tappolet, C., Stroud, S. (eds), 2003. 17-38

- ------, 2004/2002. "Evaluation, Uncertainty, and Motivation", originally in <u>Ethical Theory and Moral Practice</u>, 5, pp. 305-320

- ------, 2004/2001. "The Incoherence Argument; A Reply to Shafer-Landau", originally in Analysis, 61, pp. 254-266

- ------, 2000. "Moral Realism", originally in LaFollete, H. (ed), 2000, pp. 15-37.

- ------, 2004/1998. "Ethics and the A Priori; A Modern Parable", originally in *Philosophical Studies*, 92, pp. 149-174,

- ------, 2004/1997, "In Defence of The Moral Problem: A Reply to Brink, Copp and Sayre-McCord", originally in <u>Ethics</u>, 108, pp. 84-119.

- ------, 2004/1995, "Internal Reasons", originally in <u>Philosophy and Phenomenological Research</u>, 55, pp. 109-131.

- ------, 2004/1995b. "Internalism' Wheel", originally in <u>Ratio</u>, 8, pp. 277-302

- ------ 1994, *The Moral Problem*, Blackwell, Oxford.

- ------, 1994a. "Why Expressivists about Value should Love Minimalism about Truth?", <u>Analysis</u>, 54, pp. 1-11

- ------, 2004/1988, "On Humeans, Anti-Humeans, and Motivation. A Reply to Pettit", originally in <u>Mind</u>, 97, pp. 589-595

- ------, 1987. "The Humean Theory of Motivation", <u>Mind</u>, 96, pp. 36-61

- Soames, S. 2003. *Philosophical Analysis in the Twenty Century. Vol. 1. The Dawn of Analysis*, Princeton, Princeton University Press.

- Stevenson, C. L. 1944. *Ethics and Language*, New Heaven, Yale University Press

- Stich, S. Doris, J. "Empirical Perspectives on Ethics", in Jackson, F. Smith, M. (eds). 2005. pp. 114-152.

- Strawson, P. 2003/1962. "Freedom and Resentment", in Watson, G. (ed). 2003, originally in Proceedings of the British Academy, 48, pp. 1-25

- Sturgeon, N. 2006. "Moral Explanations Defended" in Dreier, J. (ed). 2006

- ------, 2006a. "Ethical Naturalism", in Copp, D. (ed). 2006.

- ------, 2003. "Moore on Ethical Naturalism", Ethics, 113, pp. 528-556

- ------, 1998/1985. "Moral Explanations", in Rachels, J. (ed). 1998, originally in Copp, D. Zimmerman, D. (eds). *Morality, Reason, and Truth*, Rowman & Littlefield, pp. 49-78

- Svavarsdóttir, S. 2006. "How Do Moral Judgments Motivate?", in Dreier, J. (ed). 2006. pp. 163-181

- ------, 1999. "Moral Cognitivism and Motivation", The Philosophical Review, 108, pp. 161-219

- Tenenbaum, S., 2006. "Directions of Fit and Motivational Cognitivism", in Shafer-Landau, R. (ed), 2006, 235-265

- Thompson, J. J. 2007. "Normativity", in Shafer-Landau, R. (ed). pp. 240-267

- ------. 1996. "Moral Objectivity" in Harman, G. Thompson, J. J. 1996.

- Timmons, M. 1998. *Morality without Foundations. A Defence of Ethical Contextualism*, Oxford, Oxford University Press

- van Roojen, M. 2004. "Moral Cognitivism vs. Non-cognitivism" in Eward N. Zalta (ed), *The Stanford Encyclopedia of Philosophy*, Winter 2004 Edition,

- Velleman, J. D, 2000, *The Possibility of Practical Reason*, Oxford, Oxford University Press.

References

- ------, 2000a, "Introduction to *The Possibility of Practical Reason*", in Velleman, J. D., 2000. 1-31

- ------, 2000b. "On the Aim of the Belief" in Velleman, J. D. 2000. Chapter 11.

- ------, 2000/1996, "The Possibility of Practical Reason", originally in <u>Ethics</u>, 106, 694-726

- ------, 2000/1992, "The Guise of the Good", originally in <u>Noûs</u>, 26, 3-26.

- ------, 2000/1992a, "What Happens When Someone Acts?", originally in <u>Mind</u>, 101, 461-481

- ------, 2000/1989. "Epistemic Freedom", originally in <u>Pacific Philosophical Quarterly</u>, 70, pp- 73-97

- Wallace, R. J. (forthcoming). "The Deontic Structure of Morality"

- ------, 2006, *Normativity & the Will. Selected Essays on Moral Psychology and Practical Reason,* Oxford, Oxford University Press.

- ------, 2006a, "Moral Reasons and Moral Fetishes: Rationalists and Anti-Rationalist on Moral Motivation", in Wallace, R, J., 2006, 322-342.

- ------, 2006b. "Moral Motivation" in Dreier, J. (ed). 2006. pp. 182-196

- ------, 2005, "Moral Psychology", in Jackson, F., Smith, M., 2005, pp. 86-114.

- ------, 2006/2004. "Normativity and the Will", originally in Hyman, J. Steward, H. (eds). <u>Agency and Action</u>, Cambridge, Cambridge University Press, pp. 195-216

- ------, 2006/2003, "Explanation, Deliberation, and Reasons", originally in <u>Philosophy and Phenomenological Research</u>, 67, 429-435.

- ------, 2006/2002, "Scanlon's Contractualism", originally in <u>Ethics</u>, 112, (April), 429-470.

- ------, 2006/2001, "Normativity, Commitment, and Instrumental Reason", originally in <u>Philosophers' Imprint</u>, 1/3, December, http://hdl.handle.net/2027/spo.3521354.0001.004, reprinted in Wallace, R. J. pp. 82-120.

- ------, 2006/2000. "Caring, Reflexivity, and the Structure of Volition", pp. 190-211, originally in Beltzer, M. Guckes, B. (eds). *Autonomes Handlen*, Berlin, Akademie Verlag, pp. 213-234

- ------, 2006/1999. "Three Conceptions of Rational Agency", originally in <u>Ethical Theory and Moral Practice</u>, 2, pp. 217-242

- ------, 2006/1990, "How to Argue about Practical Reason", originally in <u>Mind</u>, 99, (July), 335-385

- Wallace, R. J. Pettit, P. Scheffler, S. Smith, M. (eds). 2004. *Reasons and Value. Themes from the Moral Philosophy of Joseph Raz*, Oxford, Oxford University Press.

- Watson, G. 2004. *Agency and Answerability*, Oxford, Oxford University Press

- ------, 2004/2003. "The Work of a Will", originally in Stroud, S. Tapollet, C. (eds). *Weakness of Will and Practical Irrationality*, Oxford, Oxford University Press.

- ------, (ed). 2003. *Free Will*, Oxford, Oxford, University Press

- ------, 2004/1977. "Skepticism about Weakness of Will", originally in <u>Philosophical Review</u>, 86, pp. 316-339

- ------, 2004/1975. "Free Agency", originally in <u>Journal of Philosophy</u>, 78/2, pp. 205-220.

- Wedgwood, R. 2007. *The Nature of Normativity*, Oxford, Oxford University Press

- Williams, B., 1995, "Internal Reasons and the Obscurity of Blame", in Williams, B., 1995, *Making Sense of Humanity and Other Philosophical Papers: 1982-1993*, Cambridge, Cambridge University Press

- ------, 1981, "Internal and External Reasons" in Williams, B., 1981, *Moral Luck*, Cambridge University Press, Cambridge.

- ------, 1971. "Deciding to Believe", reprinted in Williams, B. 1973. *Problems of the Self*, Cambridge, Cambridge University Press

- Wright, C. 1992. *Truth and Objectivity,* Cambridge, Harvard University Press

- ------, 1987. "Realisms, Antirealism, Irrealism, Quasi-Realism", <u>Midwest Studies in Philosophy 12</u>

- Zangwill, N. 2006. "Moral Epistemology and the Because Constraint", in Dreier, J. 2006.

References