
UNIVERSIDAD
DE GRANADA

**Programa de Doctorado en Psicología**

**Departamento de Psicología Experimental**

**Tesis Doctoral**

# Deliberation and Morality

**Doctorando Daniel Frances Spears**

**Directoras**

**Felisa González Reyes y María Ruz Cámara**

**2021**

Editor: Universidad de Granada. Tesis Doctorales
Autor: Daniel Frances Spears
ISBN: 978-84-1306-970-8
URI: http://hdl.handle.net/10481/69870

# Acknowledgments and Dedications

As much work and time has gone into this thesis, I would like to take a moment to acknowledge a few people who have been there and helped me along the way to completing this. It has not been easy for me, my friends, my directors, and especially my family. First, I would like to thank my directors, Felisa González and María Ruz for their endless patience, understanding, and I would like to point out that I know it could not have been easy working with a part time student. I must also thank Yasmina Okan for her endless help, advice, and time she dedicated to further my thesis, though she did not have to. Thank you to José César Perales, Inés Fernández Linsenbarth, and Irene Hinojosa Aguayo for their help on my articles and research. Importantly, as well, I would like to thank Jonathan Contreras Bustos for employing me all these years, being flexible with my teaching hours, and always being available to help when I needed it. Further, I must acknowledge my friends and colleagues: Itsaso Colás, Javier Ortiz, Julian Hernández, Carlos González, Maïka Telega, Paloma Díaz, Ana Paqui Polenciano, Alberto Sobrado, y Juan Eloy Arco who helped me uncountable times for uncountable reasons. Without them, I may not have been able to achieve this. My friends, Kyle Omholt, Jon Zeitler, Kurt Neubauer, Tony Herriot, Jamie McVey, Daniel Eriç, Peter Walstow, Bethany Slawson, Juliet Rowe, Peter Weller, Laura Hansen, and Tony and Chessa Bruce, who have listened for years as I explained how I was ‘almost finished’. Finally, I would like to acknowledge my Granada ‘godparents’, Miles and Sue McCarthy who have treated me like family, and giving me a place to stay for Christmases and any holiday needed; and Matthew and Ana Poë, who brought me in and helped guide me for years. Finally, thank you to my uncle Matthew Fitzgibbon, who started the trail into academia for me to follow. Thank you all so much. You will never know how much your support and understanding has meant to me.

This work is dedicated to my brothers who struggled to understand my research. To my mother who has waited for me to finish, and visit. To my father who supported me, but sadly was not able to be here to see me finish (May he rest in peace). Finally, this is dedicated to my fiancée, Emma Kane, who has waited for me so we can move back Ireland.

# ÍNDICE DE CONTENIDOS

## Resumen

Desde la perspectiva de la Teoría de Proceso Dual de la Moralidad (Greene et al., 2001, 2004), en la que se utilizan dilemas morales sacrificiales (ej. una persona tiene que decidir si sacrificar o no una vida para salvar otras cinco), se considera que las decisiones utilitarias (sacrificar) son deliberativas, mientas que las deontológicas (no sacrificar) serían intuitivas. Aunque esta aproximación teórica no está libre de críticas, existe evidencia que apoya la conexión utilitarismo-deliberación. Los objetivos principales de esta tesis son a) ampliar nuestro conocimiento sobre la conexión entre la toma de decisiones utilitarias en el ámbito moral y el pensamiento deliberativo, incluyendo la posibilidad de utilizar manipulaciones experimentales que permitan inducir pensamiento deliberativo que podría afectar a la toma de decisiones moral, haciéndola más utilitaria; b) estimar el grado de relación entre diferencias individuales en rasgos de reflexión cognitiva y psicopatía ('blunted emotion' aplanamiento afectivo) con el utilitarismo mostrado en dilemas morales; así como c) estudiar cómo deliberación y moralidad están conectadas con decisiones y actitudes en situaciones del mundo real.

*Primer objetivo: inducción de pensamiento deliberativo*

El primer objetivo se centra en la posibilidad de inducir un tipo de procesamiento deliberativo para, posteriormente, comprobar si se produce un incremento de las decisiones utilitarias en dilemas morales. Para cumplir este objetivo, se utilizaron tres tipos de manipulaciones distintas.

En el primer estudio (Capitulo 3, Estudio 1), basado en el experimento de Paxton et al. (2012), los participantes completaron bien el *Cognitive Reflection Test* (CRT, Frederick, 2005), bien el *Berlin Numeracy Test* (BNT; Cokely et al., 2012) antes de contestar una serie de dilemas

morales. Además, y a diferencia del estudio de Patxon y colaboradores, dos de los grupos recibieron feedback sobre sus respuestas incorrectas, una pista para corregir el error y una nueva oportunidad de contestar (grupos feedback), como una manipulación adicional de inducción de un estilo deliberativo de pensamiento, mientras que otros dos grupos no recibieron este tratamiento. Hubo por tanto cuatro grupos: CRT y BNT (sin feedback) y CRT_fb y BNT_fb (con feedback), bajo la hipótesis de que los grupos que recibieron feedback verían incrementada la probabilidad de utilizar un estilo de pensamiento deliberativo y, por tanto, mostrarían decisiones más utilitarias. Además, una segunda hipótesis inicial era que el efecto de inducción de estilo deliberativo sería mayor en el caso del grupo CRT, puesto que se supone que, además de cognición numérica que requiere un procesamiento deliberativo, la tarea requiere la inhibición de respuestas preponderantes incorrectas. Este proceso de inhibición, aplicado a las respuestas emocionales o intuitivas inducidas por el contenido de los dilemas morales, se relaciona con opciones de respuestas más utilitarias. Sin embargo, los resultados mostraron que, aunque el feedback incrementó las respuestas utilitarias, el efecto se debió fundamentalmente al grupo BNT_fb, no al CRT_fb.

Una segunda manipulación para inducir pensamiento deliberativo (Capitulo 3, Estudio 2) fue utilizar un tipo de letra difícil de leer (disfluencia perceptiva). Alter et al. (2007) encontraron que la presentación de los ítems del CRT escritos con una fuente de letra difícil de leer (disfluent), suponía una mejora significativa en la ejecución en comparación con una fuente de letra normal o fácil de leer, lo que podría relacionarse con la inducción de un estilo de pensamiento más deliberativo. Dado que la ejecución en CRT correlaciona con las repuestas utilitarias a dilemas morales, la hipótesis de partida era que la utilización de una fuente de letra *disfluent* podría inducir deliberación y, por tanto, respuestas más utilitarias a los dilemas morales. También, dado que las

respuestas utilitarias a dilemas impersonales (donde no se inflige un daño directo a la víctima a sacrificar) son más altas, generalmente, comparadas con los dilemas personales (donde sí hay un mayor sentido de agencia y por tanto más repuestas deontológicas de 'no sacrificar'), se predice que el efecto de *disfluency* será mayor en el caso de dilemas personales comparados con dilemas impersonales. En este estudio se utilizó un diseño intrasujeto con los factores disfluency y tipo de dilema, contrabalanceándose el orden de presentación de los bloques correspondientes al primer factor (disfluent or fluent). Los datos revelaron que, cuando los participantes leían dilemas en una fuente *disfluent*, tomaron más decisiones utilitarias comparado a cuando los leyeron en una fuente normal, si bien se observó un efecto de orden: el efecto ocurría cuando el bloque de dilemas *difluent* se presentaba en primer lugar. No hubo ningún efecto del factor tipo de dilemas.

El tercer estudio experimental (Capitulo 3, Estudio 3) consistió en dos experimentos donde se usaron tareas de control cognitivo para inducir un estilo deliberativo con el fin, de nuevo, de constatar si se producía un incremento paralelo en decisiones utilitarias. Green et al. (2001, 2004) compararon las áreas del cerebro que activaban cuando una persona tomaba decisiones utilitarias con aquellas implicadas al contestar ensayos incongruentes en una tarea Stroop (Stroop, 1935). En el primer experimento, los participantes contestaron dos tipos de bloque: bloques con un 50% de ensayos incongruentes (alta carga) y bloques con 100% de ensayos congruentes (baja carga), contrabalanceado el orden por tipo de bloque. Después de cada bloque, contestaron a un dilema moral. Hubo 3 bloques de alto conflicto, y 3 bloques de bajo conflicto, por lo que los participantes contestaron a 6 dilemas en total. En el segundo experimento se utilizó el mismo diseño con una tarea de flancos en lugar de una Stroop. Se esperaba que los participantes mostraran más decisiones utilitarias después de bloques de alta carga comparado con los de baja carga. En el experimento con la tarea Stroop no hubo efecto de carga sobre las decisiones morales. Sin embargo, en el

experimento de flancos se encontró que las decisiones utilitarias disminuían después de completar bloques de alta carga en comparación con las decisiones tomadas después de completar bloques de bajo carga.

*Segundo objetivo: deliberación, utilitarismo y psicopatía*

Un segundo objetivo de la tesis (Capitulo 4, Estudio 4) era estudiar las diferencias en las relaciones entre deliberación y utilitarismo, y psicopatía y utilitarismo en el ámbito moral. Bartels et al. (2011) encontraron que una puntuación mayor en rasgos de psicopatía se relacionaba con puntaciones utilitarias en dilemas morales, presuntamente por una diminución en emociones sociales. Dado que también existe una relación entre estilo de pensamiento deliberativo y utilitarismo, una pregunta que se plantea en esta tesis es la siguiente. ¿Hay relación entre decisiones utilitarias, psicopatía, y deliberación? Para responderla, en este estudio un grupo de participantes completó un conjunto de dilemas morales, un cuestionario de psicopatía (SRP-III), y el CRT. Se encontró que, mientras que tanto las respuestas correctas en el CRT y las puntuaciones en psicopatía correlacionaron positivamente con las respuestas utilitarias a dilemas morales, no hubo sin embargo una correlación significativa entre ejecución en el CRT y puntuación en la escala de psicopatía.

*Tercer objetivo: deliberación, utilitarismo, fundamentos morales y actitudes sociopolíticas*

El tercer objetivo (Capitulo 5, Estudio 5) de la tesis era ampliar nuestra comprensión sobre la relación entre moralidad y pensamiento deliberativo en conexión con actitudes o creencias políticas y económicas. Las decisiones deontológicas se han asociado con la ideología conservadora (Antonenko Young et al., 2013; Piazza & Sousa, 2014), aunque podría afectar al conservadurismo social y no al fiscal o económico (Chan, 2019). También, ciertos fundamentos

morales (Pureza, Autoridad, y Lealtad; Graham et al., 2011) muestran relaciones negativas con respuestas utilitarias. Además, se han reportado relaciones negativas significativas entre medidas de conservadurismo y medidas de deliberación (Chan, 2019; Deppe et al., 2015).

Sin embargo, en estos estudios las medidas de orientación política eran muy sencillas (usualmente un único ítem) y solamente utilizaban el CRT para medir pensamiento deliberativo. En el estudio incluido en esta tesis se utilizan dos medidas para aproximarse al estudio de la moralidad (dilemas morales, y el MFQ, un cuestionario para estimar el grado de apoyo a distintos fundamentos morales); tres medidas de orientación y actitudes políticas/económicas (ítem de orientación política; Escala de Conservadurismo Social-Económico, SECS, con sus dos subescalas, social y económica/fiscal; y Apoyo al Mercado Libre); y tres medidas de deliberación (CRT, BNT, y BBS, una tarea de razonamiento silogístico, para medir el sesgo de creencia, o *Belief Bias*).

Las hipótesis de partida principales fueron que las decisiones utilitarias correlacionarían negativamente con las medidas conservadurismo; que las decisiones utilitarias correlacionarían negativamente con los fundamentos morales asociados al conservadurismo político (Autoridad, Lealtad y Pureza); y que las medias de conservadurismo correlacionarían negativamente con las medidas de pensamiento deliberativo. Se encontró que, para algunos dilemas, las respuestas utilitarias correlacionaron negativamente con medias conservadurismo; asimismo, algunos fundamentos morales relacionados con una postura liberal correlacionaron positivamente con repuestas utilitarias a dilemas morales (Justicia), mientras que otros, asociados al conservadurismo, correlacionaron negativamente (Pureza). Respecto a medidas de pensamiento deliberativo y de orientación económica-política, se encontró que la ejecución en el BBS correlacionó negativamente con todas las medidas de conservadurismo, encontrándose también

relaciones negativas entre ejecución en el CRT y medidas de conservadurismo. Además, los fundamentos morales de Autoridad y Pureza (asociados a la orientación conservadora) mostraron correlaciones significativas negativas con todas las medidas de pensamiento deliberativo (CRT, BNT, y BBS), mientras que la puntuación en Lealtad (conservadurismo) correlacionó negativamente con la puntuación en BBS.

Los resultados incluidos en esta tesis tienen implicaciones teóricas y en el mundo real. Respecto al primer objetivo (Capitulo 3), donde se utilizaban manipulaciones experimentales con el fin de inducir pensamiento deliberativo, los resultados del primer experimento (Capitulo 3, Estudio 1) indican que las decisiones utilitarias no solo están asociadas con cognición deliberativa (Greene et al., 2001, 2004; Paxton et al., 2012), sino que la necesidad de deliberar más en una tarea numérica (feedback) puede incrementar paralelamente el grado de pensamiento deliberativo en comparación a responder a la tarea sin feedback, lo que, a su vez, parece incrementar las respuestas utilitarias a dilemas morales. Junto con los resultados del experimento de disfluencia (Capítulo 3, Estudio 2), donde se observa también que una manipulación asociada a una mejor ejecución en tareas deliberativas incrementa a su vez las respuestas utilitarias (al menos cuando el bloque de dilemas *disfluent* se presenta en primer lugar), la efectividad de las manipulaciones que inducen pensamiento deliberativo para incrementar las respuestas utilitarias apoyarían la Teoría de Proceso Dual de Moralidad (Greene et al., 2001, 2004). Sin embargo, los resultados no son tan claros en el caso del tercer estudio (Capitulo 3, Estudio 3) donde se utilizaron tareas de control cognitivo (Stroop y flancos). El hecho de que completar bloques de alto conflicto reduzca las respuestas utilitarias a dilemas morales podría ser congruente con la teoría del proceso dual si tenemos en cuenta que una carga cognitiva demasiado alta puede, de hecho, disminuir las respuestas utilitarias

(Timmons et al., 2018; Témolière et al. 2012), en lugar de aumentarlas, dado que consumiría recursos cognitivos e interferiría con otra tarea que también los requiera.

También en esta tesis, llama la atención la respuesta a la cuestión de si las decisiones utilitarias son 'frías' o deliberativas. Bartels et al. (2011) y Kahane et al. (2012) establecieron una conexión entre psicopatía y decisiones utilitarias. En relación a esto, la cuestión que consideramos en esta tesis es si las personas que son más deliberativas y eligen opciones utilitarias también muestran rasgos más psicopáticos o de frialdad emocional en al ámbito social. En el Capítulo 4, los resultados muestran que, aunque las respuestas utilitarias correlacionan con respuestas correctas en el CRT y con puntuaciones altas en psicopatía, no se encontró correlación significativa entre la puntuación en psicopatía medida a través del cuestionario y respuestas utilitarias a dilemas morales. Esto podría implicar la existencia de dos 'vías' distintas, relacionadas ambas con las respuestas utilitarias a dilemas morales, pero no relacionadas entre sí, como serían la tendencia a pensar de una manera más deliberada, calculando el mayor bien o beneficio, mientras que la otra involucraría una frialdad emocional por la que no se mostraría la aversión a dañar a otros que normalmente podría oponerse a la opción utilitaria.

La conexión entre las decisiones utilitarias en los dilemas morales y los fundamentos morales también es importante tanto para la Teoría de Proceso Dual de Moralidad como para la MFT (Moral Foundations Theory). Crone et al. (2015) encontraron que las decisiones utilitarias se relacionaban negativamente con los fundamentos morales de Cuidado y Pureza. Sin embargo, en esta tesis (Capítulo 5, Estudio 5) se encontró que las respuestas utilitarias a algunos dilemas se correlacionaron positivamente con Cuidado y Justicia (fundamentos liberales), mientras que otros correlacionaron negativamente con Pureza (fundamento conservador). Por lo tanto, la conexión

entre los fundamentos morales y utilitarismo puede depender del dilema utilizado y la conexión entre la Teoría del Proceso Dual de la Moralidad y la MFT podría ser menos robusta de lo pensado.

Hay varias implicaciones de los resultados de la tesis en el mundo real también. La conexión entre la política y moralidad parece ser considerable y se ha planteado desde hace tiempo. John Stuart Mill, por ejemplo, reflexionó sobre utilitarismo, justicia, y leyes (Mill, 1863). También existe una conexión entre moralidad y filosofía económica (Crespo, 1998). Las correlaciones negativas entre conservadurismo y respuestas correctas en el BBS sugieren que es posible que las personas de orientación conservadora sean más susceptibles a sesgos cognitivos, como el sesgo de creencia, siendo más proclives al razonamiento intuitivo basado en expectativas y creencias previas, en lugar de llegar a una conclusión válida a través de los principios de la lógica.

La comprensión de la relación entre deliberación y utilitarismo es necesaria para incrementar nuestro conocimiento, no sólo acerca de la moralidad y los procesos psicológicos implicados, sino también del pensamiento deliberativo y de cómo éste afecta a cuestiones en contextos reales. En esta tesis se señala no solo cómo la deliberación y su inducción afecta a la toma de decisiones morales a través de la utilización de dilemas morales hipotéticos, sino también cómo la deliberación está implicada en el ámbito moral de la política y la economía, áreas que afectan a todos en contextos diferentes al del laboratorio

Asimismo, el efecto de la inducción del pensamiento deliberativo sobre las decisiones utilitarias, incrementándolas, no sólo muestra que la cognición moral puede verse afectada por el pensamiento deliberativo, sino que éste puede ser inducido. Sin embargo, la efectividad de la inducción puede estar limitada por factores como la dificultad de las tareas cognitivas o la carga cognitiva que requieren.

**Forward**

Imagine, for a moment, that you are living in a village outside Palmyra in Syria. One day, a group of Islamic State of Iraq and Levant (ISIL) militants come through that village on their way to Palmyra. You have heard stories about them killing everyone and destroying everything in their path. Upon hearing that the soldiers are nearing your village, you, your spouse, your neighbors, and their baby seek refuge in a hidden compartment under your house. You are the first to climb into the refuge, and therefore must take the baby due to the design of the hiding space.

The entrance is well hidden from sight, but any noise can be heard clearly, so it is important that everyone remains extremely quiet when the militants come. As the minutes pass, the voices of the soldiers grow louder, and you can hear them going in and out of houses, destroying belongings and killing any villager they find. There is no enemy fire against them, so once they are finished going through your village, you all will be able to escape to safety. The soldiers burst into your house and begin destroying things. You can hear them throwing things on the floor, smashing glasses, smashing the windows, and busting the walls. After several minutes, the boots start to shuffle around. They are going to be leaving soon. You are almost in the clear. You all are going to live!

Then the baby begins to cry…

The first cry is soft, and you can tell the soldiers did not hear it, but you put your hand over its mouth. The cries are getting stronger, and your hand is barely muffling the sounds. The boot steps of the soldiers have stopped moving, and become silent, listening. If the soldiers hear baby's cries, they will find the hiding space and kill everyone, including the baby. You notice that your hand is so hard over the mouth that it is beginning to asphyxiate the baby. You must decide now:

remove your hand, letting the baby cry out, and alerting the soldiers to where you are, which will get you all killed; or kill the baby to save everyone else hidden. There are no other choices. There are no loopholes to save everyone. This scenario, basically, is a representation of the types of dilemmas used in Moral Decision-Making.

Neither choice is simple or easy because both options will bring about death. This elaborated example is a version of a moral dilemma typically used in research on moral decision-making (Baron, Scott, Fincher, & Metz, 2015; Greene, Nystrom, Engell, Darley, & Cohen., 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Paxton, Ungar, & Greene, 2012) which were originally part of a philosophical thought experiment, famously titled 'the Trolley Problem' that pit two different philosophical moral arguments against each other (Foot, 1967). These early studies found evidence that decisions to save more people while sacrificing another person are associated with rational or deliberative cognition (Baron et al., 2015; Greene et al., 2004; Greene et al., 2001; & Paxton et al., 2012). What, then, leads people make these choices?

Using dilemmas like these, this thesis attempted to increase our understanding of moral decision-making by examining the effect of the induction of changes in participants' cognition (i.e., trying to prompt deliberative cognition) to determine if these increase the likelihood of sacrificial choices to save more lives, which is considered the normative choice (i.e., utilitarian), further investigating the connection between deliberation and utilitarian moral decisions on these types of dilemmas (i.e., trying to induce deliberative thinking). In addition, although some literature connects the utilitarian sacrificial aspects of decisions making in dilemmas with blunted emotions (Bartels & Pizarro, 2011), there is a concern regarding whether these 'cold and calculating' decisions are necessarily rational or not, therefore this thesis aimed at investigating whether measures of callous affect relates to those of reflective thinking. Finally, to increase the

relevance of this thesis to real-world issues, this thesis also uses other tasks other than sacrificial moral- dilemmas scenarios to attempt to broaden our understanding of the relationship between morality, politics, and deliberative cognition, in a broader examination that has previously been attempted by using more exact measures of political and economic ideologies.

# Part I: Introduction

# Chapter I

# Introduction

## 1.1. Definitions and introductory concepts

What is a moral dilemma? What are they good for? The first use of moral dilemmas in Psychology did not begin until 2001, by Joshua Greene (Greene et al., 2001). As he explains in his book *Moral Tribes* (Greene, 2014), being a member of debate teams, Greene would often use utilitarian arguments to beat his opponents. However, one day during a debate, he was confronted with a version of the Trolley problem where a doctor could sacrifice one healthy patient to harvest their organs in order to save five sick patients. Though he was a true believer in utilitarian morality, he could not respond. This inspiration led him not down a philosophical path, but to a psychological one, where he wanted to understand what factors affected the choices of people on such dilemmas.

### *Deontology vs. Utilitarianism*

The two decisions in the dilemma, to harm or not to harm, represent two distinct and generally opposed moral philosophical arguments. The decision to not do any harm is typically referred to as a deontological decision. Deontology is concerned with, not only the action, but also the intention of the action (Kant, 1797). Conversely, the choice to harm another person for the sake of a greater good is frequently referred to as the utilitarian decision (Greene et al., 2001) and it is often referred to as a consequentialist argument (Kahane et al., 2018) since the moral good is the outcome of an action, regardless of the action. This philosophical argument of utilitarianism states that a person should do that action which brings the least amount of aggregate harm (Mill, 1863; (Kahane et al., 2017; Manfrinati et al., 2013b). Consequentialism argues that an action is considered moral because the result is good, instead of the action itself, nor the intention of the person (Anscombe, 1958). As it will be discussed later in the thesis, some argue that the

deontological responses are immediate, based on emotion and therefore are intuitive, while utilitarian responses are slower, deliberative, and thus more rational (Greene, 2007; Greene et al., 2004; Greene et al., 2001). However, as will be seen later, some of these premises have been challenged.

*Type of moral dilemmas*

Moral dilemmas have also been commonly categorized into different types. Dilemmas where a person diverts a deadly threat from a group to an individual—similar to the classical *Trolley problem* where a person must pull a lever to switch the tracks of an out-of-control trolley to save five lives while sacrificing one on the new track—are considered *impersonal* (Greene et al., 2004). Dilemmas where a person is directly sacrificed, such as the example given in the preface, instead of diverting the threat from the group to the individual, have been referred to as *personal* dilemmas (but see McGuire et al., 2009 for a critique on the personal/impersonal distinction). There have been further distinctions made such as high and low conflict dilemmas (Koenigs et al., 2007), and self or other, inevitable or avoidable (Moore et al., 2008).

High and low conflict dilemmas are types of personal dilemmas. A high conflict dilemma has been empirically defined as one where people with ventromedial prefrontal cortex lesions make more utilitarian decisions and make them faster compared to subjects without lesions (Koenigs et al., 2007). Low conflict dilemmas are those where participants with and without lesions make the same decision without any significant time differences (usually giving non-utilitarian responses). Moore et al. (2008) separated dilemmas into ‘self’, where the actor will die if they do not sacrifice a person, or ‘other’ where the actor will not die regardless of the decision they make. The authors further separated dilemmas into inevitable, where the person who is to be

sacrificed will die regardless of what the actor does; or avoidable, where the person who would be sacrificed will survive if the actor does not kill them. Moore et al. (2008) found that people with higher working memory capacities made more utilitarian decisions on personal dilemmas where the victims' death was inevitable and in personal dilemmas in general.

Recent research has used process dissociation (PD) analysis to quantify whether decisions are pro-utilitarian, pro-deontological, or pro-harm thinking by using congruent and incongruent dilemmas (Conway & Gawronski, 2013). While these were not used in any study in this thesis, they are important in the understanding some relevant research. Congruent dilemmas are dilemmas where utilitarian decisions and deontological decisions are the same. For example, imagine that the enemy soldiers in the preface example were capturing the people in the village and putting them in labor camps where they will suffer, but not necessarily die. The decision to asphyxiate the baby is not deontological because it breaks an extreme rule to 'do no harm' but, importantly, it is not utilitarian either since the harm done compared the harm prevented is not equal or greater. That is, killing the baby to stop others from being made to do forced labor causes more aggregate harm than it does aggregate good. Those who endorse the sacrificial decisions in these congruent dilemmas are neither deontological nor utilitarian, but rather pro-harm. Incongruent dilemmas are the traditional dilemmas, like the one in the preface, where deontological and utilitarian decisions are counter to each other. Congruent and incongruent dilemmas have helped distinguish that people with antisocial personalities endorse pro-harm (Conway et al., 2018) and not utilitarianism, as previously thought (Bartels & Pizarro, 2011). Further, these dilemmas have facilitated research showing that people who score high on Need for Cognition, reduce utilitarian decisions after cognitive load instead of increasing endorsements of deontological decisions (Conway &

Gawronski, 2013), and that analytical thinking induction (Li et al., 2018) increased support for utilitarian decisions instead of reducing support for deontological responses.

### 1.2. Moral Psychology: A brief historical perspective

One of the largest questions facing moral psychology since its beginning is whether moral decisions and judgments are intuitive/emotional versus reflective/rational. Both deontological (Kant, 1785) and utilitarian (Bentham, 1789; Mill, 1863) philosophies represented logical attempts to define what actions are 'good'. In fact, in the preface to *Grounding for the Metaphysics of Morals* (Kant, 1785) the author expresses that there is a difference between the empirical investigation of morality, and what morality 'ought' to be; that these things are different; and importantly, the rules that ought to come from rational thinking and debate. Jeremy Bentham (1789) and John Stuart Mill (1863) had a similar stance, though they believed that laws and rules are only morally good if they had a rational and reasonable utility in increasing the greatest amount of happiness (or least amount of pain) whilst Kant's deontological rules were established by his categorical imperative, which in part states that moral actions treat people as rational beings with intrinsic value and not some means to an end.

In moral psychology, rationality and intuition/emotion can affect judgments and decisions, but in different ways. There has been plenty of research inducing emotions that later affect decisions (e.g., Conway & Gawronski, 2013; Strohminger et al., 2011; Valdesolo & DeSteno, 2006). However, experiments inducing rational, reflective or deliberative cognition have been sparse (Paxton et al., 2012; Li et al., 2018) and controversial (Baron et al., 2015). Before looking at this divide, it is important to have a historical understanding of moral psychological research.

### 1.2.1 Kohlberg's Developmental Morality

Kohlberg's developmental theory of morality was one of the first (Kohlberg, 1969). He argued that moral cognition changes as a person ages and the underlying cognitive abilities improve. He outlined six cross-cultural and sequential stages of moral development, divided into three levels (Kohlberg & Hersh, 1977). The first level is the *pre-conventional*. Here, children follow cultural rules but learn them through punishments and rewards, which also establish obedience to authority figures who appear to be more powerful. At the second level, *conventional*, moral behavior pleases or helps others; morality is used to judge the intentions of other people's actions and to follow orders or rules set by others to maintain social-order are good actions. Finally, in the *post-conventional*, autonomous level, judgments and moral decisions are no longer based on an authority, but on abstract reasoning. In this level, freedom and rights are valued along with accepting other people's personal values and opinions. These values and opinions are self-chosen at this level, instead of instituted by an authority. Kohlberg claims that there is a logical comprehension at this level, with utilitarian undertones as well. Specifically, Kohlberg says that, without formal moral training or studying, people will naturally follow Rule-Utilitarianism, which is basically as it sounds, doing and judging those actions that bring about the most good as a rule or a law (Kohlberg, 1973). Included in this stage, people follow the social contract as a rule, and have some idea of natural human rights.

### 1.2.2 Bounded Rationality of Morality

While Kohlberg argued that morality is developed through various stages of maturity, the bounded rationality of morality argues that moral decisions are driven by heuristics, by culture, and by the environment (Gigerenzer, 2010). While only theoretical, this approach maintains that

it is nearly impossible for a person to calculate or know the optimal solution to a moral problem in real time, and that using moral heuristics can be effective and effortless. Theoretically, Gigerenzer points out how different heuristics and environments can affect a person's moral behavior more than their rational convictions (Gigerenzer, 2010). As examples, he outlines four different types of environmental heuristics in real life: 1) Imitate your peers–where a person follows the actions and decisions of their peers; 2) Equality–the attempt to divide resources equally (i.e. a parent loving their children equally); 3) Tit-for-tat interactions with other people can be affected by whether they chose to cooperate with or be kind to you or not; 4) Default–maintaining the status quo regardless of moral position (i.e. agreeing with the idea of being an organ donor, but not being one because it is an 'opt-in'[1] policy). Each of these can be modified or affected by a change in the environment a person lives in. For instance, in the equality heuristic, a parent loving their children equally can change depending on the number of children they have or, in the default heuristic, by making organ donation an 'opt-out' policy, more people will become organ donors (Johnson & Goldstein, 2003).

### 1.2.3 Social-Intuitionist Model and Moral Foundations Theory

In response to Kohlberg's developmental theory, the Social-Intuitionist Model (SIM; Haidt (2001) argued that instead of moral judgments being rational, they are *post hoc* rationalizations where people justify their judgments and decisions after they make them. Haidt (2001) list four

[1]To 'opt-in' means that a person must willfully choose to do something, in a positive way, like choosing to join a worker's union, or, in some countries, choosing to be an organ donor, and even going to university. To 'opt-out', on the other hand, means to be in something by default, and a person has to choose to leave, for example, currently on the internet, people have to 'opt-out' of what cookies a website can collect.

processes which work in a chain that make up the SIM, with two other processes which may influence or change judgments later.

The process, based on David Hume's proposal that morality was not something that was rational but came as a 'gut response' (Hume, 1777/1960), starts with the *intuitive judgment* link, where a moral judgment is an immediate reaction to some action, deciding whether it is right or wrong, good or bad. As we will talk about later, this process can be highly influenced by emotions, particularly disgust. This is followed by the *post-hoc reasoning*, where a person then justifies their response, basically convincing themselves that their response was correct. The next two processes are quite similar. The *reasoned persuasion* process is when people talk to each other about the action which they have already judged and rationalized internally. This leads to people in social groups agreeing or disagreeing, then making a 'social rule' in the *social persuasion* process. However, Haidt (2001) does yield to the fact that some judgments may be made logically with two other processes, which can override intuition, though normally when intuition is weak (*reasoned judgment* process) or change their intuition through empathetic consideration of why an immoral action was done (*private reflection*).

To study morality in this model, participants had to rate the acceptability of moral transgressions such as 'dancing sensually with a married person but do nothing more' or 'having consensual sex with a beautiful, but cheap prostitute' (e.g., de la Viña et al., 2015; Haidt et al., 1994). Additionally, participants had to justify their judgments saying whether the action harmed anyone, how it harmed them, and whether the action bothered them (the participant). The gut reaction, where participants reported feeling sick to their stomachs, was first found to affect participants who were given post-hypnotic suggestion that specific words, when seen, would cause the highly hypnotizable participant to feel an increased feeling of disgust (Wheatley & Haidt,

2005). The authors found that when the hypnotic disgust word was included in the formulation of the moral transgression, the participant judged it more harshly, regardless of how disgusting the transgression was in and of itself, sometimes while being baffled as to why they felt that way. Further, inducing disgust in various ways (smells, an unclean room, recalling a physically disgusting experience, or watching a clip from a film showing disgusting content) affected moral judgments depending upon how sensitive a person was to how their body reacts to disgust as measured by the Private Body Consciousness questionnaire (PBC; Miller et al., 1981), which measures a person's attention to their physical states (e.g., embodied moral judgment, Schnall et al., 2008): after inducing disgust, transgressions were judged more harshly compared to inducing sadness.

While this may seem as though only disgust affects judgments, other studies found that other emotions affected different types of transgressions (de la Viña, et al., 2015; Eskine et al., 2011; Seidel & Prinz, 2013). Seidel et al. (2013) observed that induced anger through annoying sounds significantly affected dilemmas violating autonomy, such as cheating on taxes, while disgust, induced by vomiting sounds, affected violations of purity. The effect of emotional induction on the evaluation of moral transgressions seems to work mainly with negative emotions rather than high arousal (or strength the emotion is felt; de la Viña et al., 2015). To summarize, emotions can intuitively guide how moral transgressions are judged, which can be amplified when emotions, particularly negative ones, are induced (de la Viña et al., 2015; Eskine et al., 2011; Haidt, 2001; Schnall et al., 2008; and Wheatley et al., 2005). Findings of this kind led Haidt and collaborators (Graham et al., 2009; Haidt & Graham, 2007) to develop the Moral Foundations Theory.

*Moral foundations*

Haidt found that there are six moral pillars that can influence moral intuition which has been referred to under the umbrella of the Moral Foundations Theory (MFT; Haidt et al., 2007; Haidt & Joseph, 2004). According to this theory, people's moral judgments (and even political beliefs) are based on six moral foundations: (1) *Care* for others, with a negative reaction towards harming others; (2) *Authority*/respect when hierarchies are respected and those at the top of the hierarchy are expected to provide protection and supplies; (3) *Fairness*, justice, or reciprocity when people are treated the same; (4) ingroup/*Loyalty* when a person trusts and values the people they are closest to, and fears those who are not; (5) *Purity*, when a person can judge an action or a person based on the amount of disgust they feel about it or when a person's actions are not decent; and, added later, (6) Liberty, where a person should be allowed to do as they please without any moral or social interference. These foundations were later confirmed through psychometric analysis of self-reporting questionnaires, which also found a connection between different foundations and political leanings (Graham et al., 2009). In particular, they observed that liberals rated Care, Liberty, and Fairness extremely high while Loyalty, Authority, and Purity were quite low. Clifford et al., (2015) developed a set of moral vignettes where different moral foundations were violated. They found negative correlations between responses to the Moral Foundations Questionnaire (MFQ; Graham et al., 2009), which measures a person's moral foundations pattern, and responses to the vignettes when the vignette violated one of the foundations the participant valued. For instance, if the vignette violated an act of Purity, and the respondent endorsed this foundation, they found it less permissible.

Relating back to the SIM, endorsements of different moral foundations can be affected by emotions. For instance, Horberg et al., 2009 found that inducing disgust using film clips increased

negative judgments of Purity violations. Further, high scores on the foundations of Care and Purity negatively predicted utilitarian decisions on moral dilemmas, while Loyalty positively predicted them (Crone & Laham, 2015). The authors explained these results relating Care and Purity with a natural aversion to injuring others and upholding the sanctity of human life, while Loyalty would relate to what may be best for the group as a whole (i.e., the utilitarian choice). Essentially, what both the SIM and the MFQ illustrate is how powerful intuition and emotion can affect moral judgments, which can even affect political beliefs. On the other hand, Reynolds et al.(2020) found that Authority, Loyalty, and Purity, negatively correlated with correct answers on the CRT, which measures reflectivity (related to deliberative thinking disposition). Therefore, differences in intuitive/emotional vs. rational morality may be associated with the moral foundations and the moral political field.

Essentially, the MFT (Haidt et al., 2007) suggests that, not only are moral judgments intuitive, but that these intuitions come from six foundations. While this gives credence to an intuitive aspect to moral judgments and decisions, it is not the only factor affecting these types of thinking.

### 1.2.4 The Dual Process of Moral Decision-Making

The dual-process theory explains moral judgments and decisions as being affected by two different cognitive processes: intuitive/emotional or reflective/rational/deliberative. Though, instead of the rational process being *post hoc*, moral decisions can sometimes be led by rational cognition from the outset. Experiments using functional Magnet Resonance Imaging (fMRI; Greene et al., 2004; Greene et al., 2001) revealed that utilitarian decisions to moral dilemmas were associated with activation of areas of the brain such as the dorsolateral prefrontal cortex (dlPFC)

and the anterior cingulate cortex (ACC), which are related to rational thinking and cognitive control (Botvinick et al., 2001; E. K. Miller & Cohen, 2001). However, when responding to personal dilemmas (as described above), areas of the brain such as the medial frontal gyrus, posterior cingulate gyrus, and the angular gyrus all activated, which have also been associated with emotions (Greene et al., 2001). Responses to these dilemmas, either the intuitive deontological or rational utilitarian, were contextualized within the dual-process with a System 1 [2] (fast, intuitive, and emotional system) that can be overcome by a System 2 (slow, rational, and deliberative; Evans, 2007; Kahneman, 2003). A further study including activation of the ventromedial prefrontal cortex (vmPFC)—an area of the brain associated with encoding emotions (Rolls, 2005) as well as the physical response to emotions (Ongur & Price, 2000), and the amygdala, an area of the brain associated with emotional reactivity (Adolphs et al., 1994; Tranel & Damasio, 1994)—found that the amygdala was associated with an emotional aversion to harming others and a lower acceptability of utilitarian responses, with the vmPFC mediating responses between the amygdala and other regions of the brain for utilitarian decisions (Shenhav & Greene, 2014).

Not only did fMRI studies find that deontological decisions were associated with emotional areas of the brain, but behavioral research also found something similar. Choe & Min (2011) found that participants reported feeling mostly guilt and sadness after making utilitarian responses to moral dilemmas, and that trait disgust and empathy, the tendency to feel disgusted or empathic,

[2] The terms ‘System 1’ and ‘System 2’ are controversial names for the different processes in general dual-process theories (Evans & Stanovich, 2013), where the preferred terms ‘Type 1’ or ‘Type 2’ have become more common. However, in many current articles on the Dual Process Models of Morality, the terms ‘System 1’ and ‘System 2’ are still in use (Bialek & Terbeck, 2016). Therefore, for consistency with articles on moral decision making, this thesis will continue to use the terms ‘Systems’ when talking about “Types” of cognition.

were associated with fewer utilitarian decisions. Ugazio et al. (2012) induced disgust which lowered utilitarian decisions while, on the contrary, inducing anger increased them. Alternatively, inducing a positive emotion or state increases utilitarian responses (Gawronski et al., 2018; Valdesolo & DeSteno, 2006), probably due to positive emotions counteracting the negative emotions triggered by the dilemmas themselves, though some positive emotions (such as mirth and elevation) can have the opposite effect (Strohminger et al., 2011) . Finally, Lee and Gino (2015) used both emotional reappraisal and emotional suppression to reduce deontological decisions, indicating that these types of decisions are emotionally based. Essentially, dilemmas cause emotional reactions (Choe et al., 2011), affecting decisions based on the type of emotion (Ugazio et al., 2012). Induced emotions may act in such a way to counteract the emotions caused by the dilemmas (Gawronski et al., 2018; and Valdesolo et al., 2006), and this is also true with the induction of emotion regulation (reappraisal or suppression of emotions), reducing deontological responses (Lee et al., 2015).

While those experiments add support for emotional responses being related to deontological decisions, fewer behavioral experiments have been done to add evidence that rational decisions can be affected through experimental manipulation (Greene et al., 2008; Paxton et al., 2012). More recently, studies inducing cognitive fatigue caused by doing a mentally challenging task have reported reduced utilitarian decisions (Timmons & Byrne, 2019), adding evidence that cognitive effort is important to make this kind of decision. Further, cueing analytical thinking by having participants read an article that expressed how the best decisions are made through thinking analytically, led participants to make more utilitarian decisions compared to those who read an article advocating intuition (Li et al., 2018). Moreover, the authors used process dissociation analysis with congruent and incongruent dilemmas, to confirm that this manipulation

increased analytical thinking styles which indeed increased utilitarian decisions, instead of decreasing deontological responses[3] .

These two series of experiments (Li et al., 2018; Timmons et al., 2019) contribute to the support that aligns utilitarian decisions with the System 2 processing since they take cognitive effort and analytical thinking. However, there are several criticisms of the dual-process theories of morality, which consider different dual-process models to explain these data. Before examining the criticisms, it will be important to have a rudimentary idea about these models.

Currently there are three main models being tested to better understand the intuitive/emotional and deliberative/rational aspects of moral cognition. However, it is important to point out that none of the objectives in this thesis is in favor of or against any of these models. Instead, it is important to understand each in terms of how they interpret the deliberative-utilitarian connection, and how the model might work when rationality is induced. Therefore, the following will be a brief explanation of each model, rather than a detailed analysis.

The *default-interventionist model* (Evans & Stanovich, 2013; Stanovich, 1999), applied to moral psychology (Greene et al., 2004; Greene et al., 2001), considers that deontological decisions are the product of the System 1 which can (or not) be later followed by System 2, in a serial fashion. In fact, according to this model, deontological responders do not even consider the utilitarian responses, while utilitarian responders will first activate System 1 but, afterwards, they will activate System 2 engaging deliberation thus leading to utilitarian decisions (Gürçay & Baron, 2017).

[3]Process dissociation, as explained on pages 31, is the use of congruent and incongruent dilemmas to establish whether decisions support either: 1) deontological thinking; 2) utilitarian thinking; or 3) neither.

The *parallel model* (Baron et al. 2015; Baron & Gürçay, 2017; Gürçay & Baron, 2017) argues that System 2 would work in parallel and in competition with System 1, instead of System 2 overriding System1. In the end, whichever system activation is most salient would 'win' and control the decision process. In this model, deontological responders would at least consider utilitarian responses, but not enough to override the action of System 1 (Gürçay et al., 2017).

Finally, in the *hybrid model* (De Neys, 2012; Handley & Trippas, 2015; Pennycook et al., 2015; Trippas et al., 2017) when people are faced with dilemmas they intuitively recognize that there is a conflict and consider both options. Therefore, deontological and utilitarian decisions are both processed intuitively (in a parallel sense), as products of the action of System 1 (Bialek, & De Neys, 2017). Despite the considerations of both choices, this does not mean that people deeply deliberate on the choices, only take into account both options. However, if System 2 activates, then the deliberative response will override System 1, and the utilitarian decision will be made. This is a *hybrid* model since, like the *default-interventionist model*, utilitarian decisions are driven by System 2, but also *parallel*, since both choices are equally considered and taken into account in System 1. Regardless of the model, one key point in all three of them is that utilitarian decisions are System 2 (Type 2) cognition. However, some researchers argue that utilitarian decisions are not the production of System 2, casting doubts on the rational-utilitarian connection.

### *1.3 Criticisms to the dual process model of morality*

Two of the biggest criticisms of the dual-process models of morality are that: 1) utilitarian decisions on moral dilemmas are not necessarily rational decisions; and 2) sacrificial responses on popular trolley-style dilemmas do not represent utilitarianism. Firstly, some research argues that utilitarian decisions can be intuitive (Kahane et al., 2012), coming from reduced emotions rather

than rational thinking (Bartels et al., 2011; Wiech et al., 2013). Secondly, as will be discussed in more detail later, Kahane et al., (2015, 2018) point out that there are aspects of utilitarianism that are not measured by these dilemmas, introducing a scale to measure utilitarianism as a whole (Oxford Utilitarianism Scale, see below). Understanding the extent to which these dilemmas represent utilitarianism or not would help to consider their utility in psychological research. Responses to these dilemmas, and research using them, help the understanding of the ordinary person's moral thinking. Importantly, if the decision to sacrifice in these moral dilemmas do not represent utilitarianism, it weakens the connection between utilitarianism and rationality. Below is a further detailed examination of both criticisms.

### *1.3.1 First Criticism: Utilitarian responses on trolley style dilemmas are not necessarily rational.*

Kahane et al. (2012) question if utilitarian decisions are always the rational, deliberative responses. They compared traditional Trolley style dilemmas or what they called 'deontological intuitive' (D_I, where deontological decisions are the intuitive response) with others that can be considered 'utilitarian intuitive' (U_I), which are considered utilitarian intuitive since the action done to prevent greater harm is generally acceptable, according to the authors. Simply, what makes a response intuitive, U_I or D_I, is that these are immediate 'gut' responses often based in emotions or a moral rule (Greene, 2007; Haidt, 2001). In essence, the transgressions varied as well as the consequences (saving people's lives and saving people's feelings) to test whether utilitarian responses can also be intuitive. U_I dilemmas ranged from inducing an allergic reaction on a criminal who is about to go to prison in order to prevent them from spreading HIV to as many people as possible in 48 hours (*modified preventing the spread*) to lying to a friend about their spouse's affair while their marriage is going well (*white lie*). These examples are U_I because the action violates a moral rule, such as physically harming someone or lying to them for a greater

good (and because of that they are not deontological). Importantly, participants were asked to respond as soon as they had thought of their answer (which could be viewed as asking participants not to reflect, at least not too much, on their answers). The behavioral analysis revealed that participants made more intuitive decisions on both types of dilemmas, D_I and U_I. In addition, they failed to replicate Green et al.'s (2001, 2004) neuroimaging data, in the sense that they could not find activation of brain areas associated with emotions when responding to either type of dilemma, calling to doubt the emotional-intuitive connection. Essentially, their main conclusion is that the differences between utilitarian and deontological responses are based on individual differences in intuition and not on one choice being intuitive and the other being rational.

On the contrary to Kahane et al.'s (2012) findings, utilitarian decisions on traditional trolley-type D_I dilemmas have been associated with deliberative cognition. Paxton et al. (2014) used the Cognitive Reflection Test (CRT; Frederick, 2005), a numeric task of three questions where the intuitive response is incorrect, finding that participants who answered more questions correctly before responding to dilemmas made more utilitarian decisions. Questioning the findings from Kahane et al. (2012), Paxton et al., (2014) attempted to induce reflective thinking by using the CRT before participants responded to U_I and D_I dilemmas. They found a positive correlation between utilitarian choice and reflective thinking, in that the more CRT questions answered correctly, the more utilitarian decisions participants made regardless of whether the dilemmas were U_I or D_I. Further, Paxton et al. (2014) questioned the method by which utilitarian-intuitive (U_I) dilemmas were categorized since participants were asked to answer quickly, without reflecting.

Another argument stating that utilitarian decisions are not due to rational cognition is what has been referred to as 'emotional blunting'. Emotional blunting argues that, since deontological decisions are driven by emotions, utilitarian decisions come from stopping or weakening emotions

(Koenigs & Tranel, 2008). For instance, Koenigs et al. (2007) found that when responding to moral dilemmas, participants with lesions in the vmPFC, which are associated with dulled emotions and problems regulating anger and frustration (Barrash et al., 2000; Eslinger & Damasio, 1985; Shallice & Burgess, 1991) in Koenigs, & Tranel, 2007), made increased utilitarian decisions when compared to participants without lesions and participants with lesions in other areas of the brain. In addition, Bartels et al. (2011) found that high scores on self-reported Psychopathy scale (Paulhus, & Williams, 2002), a disorder associated with calloused emotions, was linked to utilitarian decisions as well. In this vein, Hayakawa et al. (2017) confirmed that, using the process dissociation procedure by reading dilemmas in a foreign language, following the study by Costa et al. (2014), increased utilitarian decision. However, the decrease in deontological thinking caused by the foreign language seemed to be due to blunting emotions rather to increasing deliberation.

In addition, several studies have related unusual patterns of increased utilitarian choices related to misbehavior, which are hardly congruent with rationality, such as reduced sense of responsibility (Franklin et al., 2009), increased levels of aggressive behavior linked to higher levels of testosterone (Carney & Mason, 2010; Montoya et al., 2013), diminished aversion to harming others associated with lower levels of serotonin (Crockett et al., 2010; Siegel & Crockett, 2013), and increased levels of drunkenness (Duke & Bègue, 2015), where dis-inhibition is associated with reduced emotional restraint (Bègue et al., 2009).

In short, the emotional blunting hypothesis claims that deliberative/rational thinking is not the only path leading to increased utilitarian decisions. Individual differences in the processing of social emotions such as guilt, shame, or empathy may also produce an unusual increase in utilitarian choices. Thus, diminishing the emotional salience of the dilemma or presenting moral

dilemmas to participants having impaired social emotion processing may well arrive to the same result: higher rate of utilitarian decisions.

Conway et al. (2018) used process dissociation and measured antisocial traits and decisions on moral dilemmas to determine the connection with utilitarian decisions and anti-social traits. They found that high scores in psychopathy (Levenson et al., 1995) and willingness to break business ethics rules (Cooper & Pullig, 2013) were related to more willingness to do harm, regardless if the outcome was utilitarian or not. When controlling for anti-social traits, this same analysis revealed that utilitarian decisions on sacrificial dilemmas were associated with empathetic concern measured by responses on the Empathetic Concern sub-scale of the Interpersonal Reactivity Index (Davis, 1983), which "measures sympathy and concern for others" (Kahane et al., 2015, p. 196). Similarly, Lee et al. (2015), using the process dissociation procedure, found that participants who regulate their emotions by suppressing their expression reduced their deontological decisions instead of increasing utilitarian ones. This is congruent with the suggestion that reduced emotion does not increase utilitarian decisions as much as it suppresses deontological ones.

### *1.3.2 Criticism 2: Utilitarian decisions on moral dilemmas do not represent true utilitarianism*

Through a series studies, Guy Kahane and collaborators (Kahane et al., 2012, 2015, 2017) challenged the interpretation that sacrificing one person to save five represents the philosophy of utilitarianism. Kahane et al. (2015, 2017) argue that true utilitarianism promotes the greater good, even if that means that a person should bring harm to themselves or those who they love by doing things, like cutting off one's own leg to save a life. Further, utilitarianism includes acts of charity, or volunteering – aspects of utilitarianism that are not captured by the dilemmas. The authors point

out that these are the two different levels to utilitarianism: *instrumental harm* (IH), which like trolley-type dilemmas uses negative actions like harming others to maximize utility (greater good), and *impartial beneficence* (IB), which is associated with giving to charity and other actions which may negatively affect the agent but have positive outcomes or benefits for more people (Kahane et al., 2017).

Kahane et al. (2015) found that utilitarian decisions on dilemmas were associated with acts or beliefs that violate IB, such as: unethical business decisions; being less charitable with donations; or not willing to self-sacrifice in order to save or help others. Since utilitarian decisions on traditional trolley dilemmas were not associated with these IB aspects, Kahane et al., (2017) created a scale which included items such as: abstract ethical questions ('from a moral perspective, people should care about the well-being of all human beings on the planet equally'); real-world questions that reflect negative utilitarian acts (doing harm, like trolley dilemmas), positive actions of utilitarianism (doing good, giving to charity); anti-utilitarian questions ('there are some things that are simply right or wrong, no matter the consequences'), as well items that illustrate criticisms or defenders of utilitarianism ('if releasing an innocent person from prison causes violent and deadly riots, then it's okay to send this innocent person to jail'). This scale, the Oxford Utilitarianism Scale (OUS), as well as its subscales (OUS-Impartial Beneficence, OUS-IB; OUS-Instrumental Harm, OUS-IH), were validated in both the lay population and ethicists (Kahane et al., 2017). The authors found that utilitarian decisions in trolley-type dilemmas positively correlated with breaking business ethical rules (IH) and negatively correlated with feeling of community, donations to charity, assisting people far away, and sacrificing oneself for the greater good, (IB). Essentially, the OUS scale, through the subscales of IB and IH, call into question

whether utilitarian decisions on moral dilemmas truly represent utilitarianism as a whole, and suggest that utilitarian decisions on moral dilemmas are more about instrumental harm.

In contrast, Conway et al. (2018) argued that there are levels to utilitarianism that define utilitarian decisions on dilemmas as a counter definition to Kahane et al. (2015, 2017) levels (IB and IH) measured by the OUS scale. The first level considers utilitarian decisions by the content of the dilemmas. They point out that, regardless of whether responses to dilemmas fulfill all of Kahane et al.'s (2017) criteria for utilitarianism, the judgments are still, by definition, utilitarian when the consequence of the sacrificial action leads to the greater good. This is true whether the person is antisocial or not. In the second level, utilitarian decisions come from a deliberative cost-benefit analyses. While this level may be cold and calculating, it still fits within the deliberative aspects of the dual-process model of morality. Importantly, as addressed above, Conway et al. (2018) found that participants with anti-social traits do not move beyond this level. The third level is when these decisions do come from a concern for the greater good (when coming from ordinary people without antisocial disorders). The fourth is a true and unfaltering commitment towards utilitarian actions in every situation, while the participant may not necessarily be a utilitarian or cognizant that they follow utilitarian ethics. Conway et al. (2017) argue that Kahane et al., (2017) begin their measurement of utilitarianism at this level, and point out that no psychologist or article has ever claimed that a utilitarian decision on sacrificial dilemma is an overall endorsement of utilitarianism. Finally, the fifth level is of people, mostly philosophers, who endorse the moral utilitarianism explicitly. The subtle difference between the fourth and fifth levels is the acknowledgment that the moral rules they follow are utilitarian.

While Kahane et al. (2015, 2017) essentially argue that utilitarian judgments start at the fourth level, Conway et al. (2018) argue that all levels are utilitarian. However, they wanted to test

whether judgments on sacrificial dilemmas were associated with concern for the greater good or not [4]. When using process dissociation, Conway et al. (2018) observed that, in the lay person, utilitarian decisions are not just a rational cost benefit analysis (second level), but also a concern for the greater good (third level). Therefore, the lay endorsement of utilitarian decisions on moral dilemmas may in fact be based on a concern for the greater good, being more than just cold and calculating decision-making. Conway et al. (2018) further defend the use of studying all these aspects of utilitarianism in psychology, including the basic first level, as it can help us to understand the cognitive underpinnings of utilitarian decisions. That is, the intentions of studying moral dilemmas are not to understand the moral cognition of moral philosophers, but what types of cognition goes into the layperson's decision making when faced with such a grim possibility of needing to sacrifice one person to save five. Furthermore, they defend these dilemmas since they can also give insight into the layperson's cost-benefit analysis when they can be affected directly by emotions, a very beneficial measurement in the real world such as with possible problems with self-driving or autonomous vehicles (Bonnefon et al., 2016; Powell et al., 2016).

In conclusion, while utilitarian decisions on moral dilemmas may not always be driven by rational, deliberative cognition (Bartels et al., 2011; Kahane et al., 2011; Koenigs et al., 2007; Lee et al., 2015), there is reason to also doubt that they only come from a lack of concern for others. Conway et al. (2018) showed that, when controlling for antisocial traits, the utilitarian decisions are indeed rational: participants who do not have antisocial traits also make these decisions as a

[4]As a note, this thesis tends to follow Conway et al.'s (2018) definition since it is not concerned with what constitutes utilitarianism psychologically-philosophically. Rather, the interest focused on under what cognitive conditions can utilitarian decisions on moral dilemmas be affected, and what types of personality traits or thinking dispositions are they associated with. However, it is important to understand that when the term 'utilitarian' is used, not all psychologists or philosophers agree that these are strictly 'utilitarian'.

response to a greater concern for the welfare of others. Philosophically speaking, utilitarian decisions on these dilemmas may not encompass utilitarian beliefs as a whole or an acceptance of utilitarianism, but there are still reasons to believe that utilitarian decisions on moral dilemmas are, at least partially, rational. As Conway et al. (2018) explained in their discussion, studying trolley dilemmas is to understand the nature of deontological and utilitarian cognition.

# Part II. Objectives

# Chapter 2

# Hypotheses and Justifications

## 2.1 Objectives & Hypotheses

As explained in the Introduction, experiments on the dual-process of morality (Greene et al., 2001) have used sacrificial moral dilemmas to investigate how emotion and cognition are associated with deontological and utilitarian decisions, respectfully, particularly when these moral philosophies diverge as they do in moral dilemmas. Although not without criticism, for the most part, research has defended a rational/deliberative-utilitarian connection vs. an emotional/intuitive-deontological link. Simply put, the central goal of this thesis is to extend the boundary of our understanding of the relation between deliberative cognition and morality, how it can be affected, and how deliberation and morality are associated with the real world, with three main objectives or goals.

The first and main objective is to attempt to induce deliberative cognition and thus increase utilitarian decisions on moral dilemmas, using different tasks and manipulations, building particularly on the research by Green et al. (2004) and Paxton et al., (2012). The second objective is to measure anti-social traits, including cold and calculating traits, and compare them with both responses to dilemmas and a measure or reflective thinking addressing the emotional blunting criticisms from Bartels et al., (2011). Finally, the link between politics, morality, and cognition, building on the moral political research in the Moral Foundations (Graham et al., 2009) and how deliberative cognition is related to the moral-political connection, was studied. Each objective is discussed in greater detail below.

1) Can rational cognition be induced thus increasing utilitarian decisions?

2) Is there a difference between rational utilitarianism and cold and calculating utilitarianism?

3) How is rational morality associated with real world morality in politics and socioeconomic attitudes?

Working Hypotheses:

### *2.1.1 First Objective:*

The first objective was to investigate whether deliberative cognition can be induced before participants respond to moral dilemmas and if this manipulation leads to a bias towards utilitarian responses.

***Hypothesis 1:*** **Induced deliberative cognition will increase utilitarian decisions on moral dilemmas.**

If utilitarian decisions are deliberative, inducing deliberative cognition using different manipulations should increase utilitarian decisions. For instance, having participants reflect further on numeric questions (CRT and Berlin Numeracy Test, BNT; Cokely et al., 2012) by providing feedback on incorrect answers should induce reflective cognition (Experiment 1). This is expected particularly in the CRT condition compared to the BNT, since correct answers on the CRT, unlike the BNT, demands overriding intuitive responses; in fact, performance on the CRT has been found to correlate with utilitarian decisions (Paxton et al. 2012). In this thesis we made use of an experimental design to deeper in the causal relationship between deliberation and utilitarianism. Further, reading dilemmas in a difficult to read font may prompt deliberative thinking (Experiment 2), according to Alter et al. (2012) who found that people answered more CRT questions correctly

when these were written in a disfluent font. This could be particularly true in the case of personal dilemmas, which are considered more emotional driven than impersonal ones. Lastly, responding to dilemmas after performing blocks of a task with a high percentage of incongruent trials which demand cognitive control (Experiment 3) should increase utilitarian decisions.

*Study. 1: Can induced reflection increase utilitarian decisions? (experimental)*

H1.a. Participants who receive feedback and some hints to reflect on their wrong response after completing a numeric task before responding to moral dilemmas will make more utilitarian decisions compared to participants who did not receive feedback.

H1.b. Participants who receive feedback on the CRT will make more utilitarian decisions compared to participants who receive feedback on the BNT.

*Study. 2: Disfluent fonts lead to more utilitarian decisions in moral dilemmas (experimental)*

H1.c. When participants read difficult to read dilemmas (disfluent fonts) they will make more utilitarian decisions compared to dilemmas written in an easier to read font.

H1.d. The effect of disfluent fonts will be greater for personal than impersonal dilemmas.

*Study 3: Can cognitive control tasks induce deliberative cognition increasing utilitarian decisions? (experimental)*

H1. e. Utilitarian decisions will increase after the participants complete blocks with more incongruent trials compared to blocks with fewer incongruent trials in a task of cognitive control.

H1.f Response times to moral dilemmas will take longer after high load blocks.

H1.g High Load blocks will have a greater affect on personal dilemmas compared to Low Load blocks

### *2.1.2 Second Objective:*

As commented in detail in the Introductions, utilitarian decisions have been associated with both deliberative cognition and blunted emotions, especially social emotions. Understanding the differential role of emotional blunting and deliberation on moral decision making, as well as their interaction, is important to increase our understanding of their contribution to utilitarian choices.

***Hypothesis 2:*** **Utilitarian decisions on moral dilemmas will be associated with reflective cognition or cold-calculating decision-making, but not both.**

We expected that:

H2.a. Participants who score high on the psychopath trait will make more utilitarian decisions.

H2.b. Participants who answer more questions correctly on the CRT will make more utilitarian decisions.

H2.c. Participants who score high on psychopath trait would not necessarily answer more questions correctly on the CRT (independent paths to utilitarian response hypothesis).

### *2.1.3 Third Objective:*

The third objective of this thesis was to examine moral decisions, deliberative thinking dispositions, and the connections with economic and political beliefs. Researchers have found relationships between deontological decisions on moral dilemmas and political conservatism (Antonenko Young et al., 2013; Piazza & Sousa, 2014), though this may be for only people with social conservative beliefs, such as the promotion of traditional marriage, and not fiscally conservative people who support the free market (Chan, 2019). In the thesis, we used other measures of morality: the Moral Foundations Questionnaire (MFQ), as it has been found that, generally, conservative moral foundations correlate negatively with utilitarian decisions to moral dilemmas (Crone & Laham, 2015; Reynolds et al., 2020).

Deppe et al. (2015) also found that conservatives answered fewer questions correctly on the CRT, which was similarly found by Chan (2019) using a self-reported cognitive style questionnaire. In this thesis, we used specific tasks to measure deliberative thinking, instead of self-reported estimation, such as numerical tasks, CRT and the BNT, as well as syllogisms to study Belief Biases, as an additional deliberative thinking disposition through logic word problems.

We also increased the number of scales to measure political attitudes to examine this objective in the thesis: a self-reported political orientation through a single-item scale, which had been used in a number of studies (Chan, 2019; Deppe et al., 2015; Lane & Sulikowski, 2017); the Social and Economic Conservatism Scale (SECS; Everett, 2013), which measures traditional values in beliefs in marriage, abortion, and crime and punishment, and fiscal conservatism, beliefs in free market ideology, social security, and business. SECS has been found to be negatively associated with correct answers on the CRT and Actively Open Minded Thinking (Yilmaz &

Saribay, 2017a); and, finally, a measure of Support for the Free Market (SFFM; Heath, & Gifford, 2006), which measures the degree to which a person agrees with a capitalist free-market. This measure has not been previously tested with moral dilemmas or measures of deliberative cognition.

***Hypothesis 3:*** **Utilitarian decisions on moral dilemmas will be negatively associated with conservative political thinking, and conservative beliefs will be negatively associated with deliberation.**

H3.a. Conservative measures in the self-reported measure, SECS and SFFM will be negatively correlated with utilitarian decisions on moral dilemmas; this will hold for social conservatives though this will not hold for economic conservatives.

H3.b. Moral foundations of Loyalty, Purity, and Authority (so-called conservative) will negatively correlate with utilitarian decisions on moral dilemmas.

H3.c. Conservative measures in the self-reported measure, SECS and SFFM will negatively correlate with correct responses on the CRT, BNT, and BBS.

H3.d. Moral foundations of Loyalty, Purity, and Authority (so-called conservative) will negatively correlate with correct responses on the CRT, BNT, and BBS.

# Part III

# Empirical Studies

# Chapter 3

# Induced Deliberation

# Study 1:

# Can Induced Reflection Affect Moral Decision-making?

Daniel Spears, Yasmina Okan, Irene Hinojosa-Aguayo, José César Perales, María Ruz, & Felisa González

Publication:

Spears, D., Okan, Y., Hinojosa-Aguayo, Perales, J.C., Ruz, M., González, F. (2021). Can induced reflection affect moral decision-making? *Philosophical Psychology*, *34:1*, 28-46. https://doi.org/10.1080/09515089.2020.1861234

## 1: Introduction

There is substantial research examining the mechanisms underlying decision making in complex moral scenarios. A common method is to present people with hypothetical dilemmas where they must decide whether they would be willing to kill one stranger to save a greater number of people. A prominent example is the *Footbridge* dilemma (Foot, 1978), where participants must indicate whether they would be willing to throw a large man over a footbridge onto rail tracks to stop a runaway train from killing five other people. The dilemma pits deontological rule-based responses ("do no harm") against utilitarian or consequentialist ones ("maximize aggregate welfare", e.g., Greene et al., 2004). The conflict relates to whether one should inflict harm on one person to reduce the aggregate amount of damage done to other people: if the large man is pushed off the bridge, the five people will be saved, but the man will have been sacrificed.

It has been argued that responses to moral dilemmas can be driven by either intuitive processes or more deliberative ones (Greene et al., 2008; Greene, et al., 2004; Greene et al., 2001). From this perspective, deciding to push the large man off the footbridge would be the deliberative choice. To make this choice, one would need to overcome the initial emotional response associated with the possibility of killing an innocent individual in such a way. Supporting this notion, Greene et al. (2001) observed that, when participants decided to sacrifice the man, areas of the brain involved in cognitive control were activated. Additionally, participants' responses were slower than when they decided not to push the man. According to this account, the decision not to sacrifice the man would be the result of fast, emotional, and intuitive processes.

This "dual process model of morality," in one form or another, has received behavioral support (Greene et al., 2008; Li et al., 2018; Paxton et al., 2012). For instance, Greene et al. (2008) manipulated cognitive load by asking participants to perform a digit search task while they

responded to moral dilemmas. Participants in the cognitive load condition exhibited longer response times for utilitarian decisions, relative to the no-load (control) condition. Other studies have found that utilitarian decisions are less likely when participants are required to complete other demanding tasks that may interfere with deliberative processes (Timmons & Byrne, 2019; Trémolière et al., 2012). Instead, utilitarian decisions can become more likely when participants are primed to deliberate either directly (Li et al., 2018), or indirectly by responding to dilemmas in a difficult to read font (Spears et al., 2018). A recent review also found that time pressure and cognitive load reduced utilitarian decisions, while individual differences in cognitive abilities, like working memory, were associated with utilitarian decisions (Trémolière, 2017). Taken together, these findings suggest that deliberative cognition is important for utilitarian decisions.

However, some studies have questioned whether utilitarian choices are always the product of deliberative processes. For instance, there is evidence that utilitarian decisions correlate with psychopathy and other anti-social personality traits not generally considered to reveal rationality (e.g., Bartels et al., 2011; Kahane et al., 2015). Both callous affect (a facet of psychopathy) and measures of cognitive reflection (Cognitive Reflection Test, CRT; Frederick, 2005) have been reported to be positively correlated with utilitarian decisions, but not with each other (Spears et al., 2014). People with anti-social traits seem to make utilitarian decisions because they are not averse to causing harm (Patil, 2015), and they would sacrifice one person to avoid others being injured (Conway et al., 2018). Relatedly, alcohol intoxication—which impairs higher order cognitive reasoning and may decrease aversion to harming others—can be related to more utilitarian decisions (Duke et al., 2015).

Other authors have also emphasized that CRT scores are not always correlated with utilitarian responses (Baron et al., 2015; Royzman et al., 2015). For example, Royzman et al.

(2015) found that CRT scores were instead related with a 'morally minimalistic' judgment pattern whereby harm-inducing acts are as seen as morally permissible but not required. That is, people with high CRT scores would deliberate about both deontological and utilitarian arguments, resulting in no behavioral preference for either action. Relatedly, Baron et al. (2015) suggested that the links between CRT scores and utilitarian responses (when found) are likely due to the existing link between CRT and the belief that it is good to question initially favored conclusions (i.e., 'open-minded thinking').

Finally, other studies have questioned the utilitarian-deliberative link by suggesting that some utilitarian decisions can be intuitive, such as not telling a friend about her husband's one-time affair to avoid causing her pain (Kahane et al., 2012). In such 'white lie' dilemmas, deliberation may be required to overcome the intuitive utilitarian response (i.e., lying) and reach the counter-intuitive *deontological* response (e.g., telling the truth). However, subsequent research found that reflective thinking (i.e., higher CRT scores) was associated with utilitarian decisions on both 'white lie' and standard footbridge dilemmas (Paxton et al., 2014).

To shed light on the role of deliberation in moral decisions, some authors have investigated specifically whether manipulations that induce reflective processes affect utilitarian responding. For instance, Paxton et al. (2012) examined the effect of answering the CRT items on people's responses to moral dilemmas. The CRT (Frederick, 2005) is thought to measure the tendency to check and inhibit prepotent intuitive responses (see Materials section, for a detailed description of the test). Responding successfully to CRT questions may make participants more reflective, leading to more utilitarian responses to moral dilemmas. Supporting this prediction, Paxton et al. (2012) found that participants who answered at least one question correctly before (but not after) responding to moral dilemmas viewed the utilitarian actions as more acceptable.

However, it is also possible that the effect of answering CRT items on people's responses to moral dilemmas found by Paxton et al. (2012) was not due specifically to increased cognitive reflection caused by the CRT, but instead merely to increased deliberation due to the requirement to engage in a numeric task prior to responding to the dilemmas. Indeed, the ability to answer numeric tasks correctly is related with deliberative thinking (Ghazal et al., 2014). Because Paxton et al. (2012) did not include a control condition with a different numeric task not involving cognitive reflection specifically, it is not possible to determine whether increased utilitarian decisions were specifically due to increased reflective cognition, or to deliberation based on numeric ability.

In the present work we aimed to shed light on the role of cognitive reflection vs. deliberation on moral decision making. Specifically, we conducted a study involving two different samples, including psychology college students and participants from the general population. Participants were allocated to complete either the CRT or a different numerical task that does not require the inhibition of intuitive responses, (i.e., the Berlin Numeracy Test; BNT; Cokely et al., 2012), and subsequently responded to a series of moral dilemmas. BNT scores are positively correlated with CRT scores, as well as with working memory span (Cokely et al., 2012; Turner & Engle, 1989). However, unlike the CRT, the BNT was not designed to lure people towards intuitive incorrect responses. That is, reaching the correct answer in the BNT does not require inhibition of preponderant intuitive responses triggered by the content of the question. Instead, correct responses on BNT items often result from increased deliberation, or "deep, elaborate processing" (Cokely et al., 2012; Cokely & Kelley, 2009) during decision-making and better self-monitoring of performance (i.e., less overconfidence; Ghazal et al., 2014). Thus, administering both the BNT and the CRT allowed us to distinguish between the effects on moral judgments of performing a

numerical task that requires deliberative processes (CRT and BNT) from the (additional) inhibition of incorrect automatic responses (CRT only). If cognitive reflection is needed to endorse utilitarian actions, we would expect participants completing the CRT to exhibit more utilitarian responses than those completing the BNT.

To distinguish between the potential effect of participants' previous reflective traits and that of performing a task that can increase reflectivity, some participants were asked to reconsider their wrong responses to either the CRT or the BNT. Specifically, participants received: a) item-by-item warning about their incorrect responses; b) a clue to solve the problem satisfactorily, and c) a chance to change their answer and respond correctly—that is, they were *further* forced to reflect (see Meyer, Spunt, & Frederick, 2015, for a similar procedure). We reasoned that providing explicit feedback on incorrect responses should encourage participants–even those who are not naturally reflective–to reconsider their responses, hence involving further reflection.

Overall, if inhibition of prepotent intuitive response is needed to make utilitarian decisions, we should find more utilitarian choices in the CRT Feedback condition than in the CRT (no Feedback) condition. If that were the case, any effect of BNT on moral choices should not, in principle, be affected by feedback. However, if the induction of deliberative thinking itself prompts utilitarian responses, feedback should increase utilitarian choices regardless of whether it is provided in connection to CRT items or to BNT items.

## 2: Method

### *2.1: Participants*

#### *2.1.1: Sample A*

A total of 248 undergraduate students from the University of Granada took part in the study in the laboratory in exchange for course credit. Following a priori exclusion criteria, we omitted 18 participants from all analyses because they declared having prior knowledge of either CRT or BNT items. Thus, the final sample consisted of 230 participants (154 female, age $M$ = 19.62, $SD$ = 2.72, range = 17 – 36).

*2.1.2: Sample B*

A total of 232 participants residing in Spain completed the study online. We omitted 10 participants from analyses because they declared having prior knowledge of CRT or BNT items. Thus, the final sample consisted of 222 participants (134 female) with a wider age range than Sample A ($M$ = 28.07, $SD$ = 10.85, range = 18 – 68). Education level was also more heterogeneous in Sample B, which included participants with no formal education ($n$ = 1), primary education ($n$ = 3), secondary education ($n$ = 74), university degree ($n$ = 90), master's ($n$ = 33), and doctorates ($n$ = 21) [5].

### *2.2: Materials and design*

All procedures were approved by the local Ethics Committee (#934/CEIH/2014). All the study materials described below, data files, and analyses scripts are available at the Open Science

[5] Of the 222 participants in Sample B, 47 were recruited through Amazon's Mechanical Turk, filtered for participants registered in Spain, and 175 were recruited through a blog hosted by the University of Granada, which was advertised in student portals and using social media. Participants recruited through the latter method included 26 undergraduate Psychology students taking part in exchange for course credit and 149 who took part without compensation. Further details on demographics for each of these subgroups can be found in Supplementary materials.

Framework (OSF): https://osf.io/8s42h/.

### *2.2.1: Tasks and feedback*

#### *2.2.1.1: Cognitive Reflection Test (CRT)*

For Sample A, we used the original CRT 3-item task (Frederick, 2005). An example item is: *'A bat and a ball cost 1.10 € in total. The bat costs 1.00 € more than the ball. How much does the ball cost?'* Respondents often provide the intuitive incorrect answer, 10 cents, instead of the correct answer, 5 cents. For Sample B, we used the new CRT version (Primi, Morsanyi, Chiesi, Donati, and Hamilton, 2016), so as to increase generalizability of our findings using a similar and equally valid task. An example item in the new version is: *'Jerry received both the 15$^{th}$ highest and the 15$^{th}$ lowest mark in the class. How many students are in the class?',* with the intuitive answer being 30, and the correct answer 29. Participants in both samples also completed a practice trial involving an additional item selected from a four-item version of the CRT (Toplak, West, & Stanovich, 2014; the "*Water Barrel*" question; see Procedure). [6]

#### *2.2.1.2: Berlin Numeracy Test (BNT)*

The BNT (Cokely et al., 2012) was used for both samples—a four-item test that measures statistical numeracy, without explicitly inducing any intuitive incorrect response. An example item is: "*Image we are throwing a 5-sided die 50 times. On average, out of these 50 throws how many*

[6] Analyses of the distribution of responses for all individual CRT (original and new) and BNT items before feedback are presented in Supplementary materials. For the CRT items, the expected incorrect answers were provided frequently, whereas other types of incorrect responses were infrequent. In contrast, for BNT items, there were no specific incorrect answers that were particularly frequent. This suggests that, as expected, only the CRT induces specific incorrect (intuitive) responses.

*times would this five-side die show an odd number?*" We used three items for the target task, whereas the remaining one (the "*Loaded Die"*) was used in the practice trial.

*2.2.1.3: Feedback*

We developed feedback statements for each item in each of the tasks described above. For example, if a participant in the CRT Feedback condition answered the *Bat and Ball* item incorrectly, they would be directed to a page displaying the following feedback: "*I'm sorry, your answer is incorrect – have you checked how much the total for the ball and bat would be according to your answer?*", with the item appearing below to be answered again. Participants in the Feedback conditions received feedback and a second chance to respond if, and only if, their response was incorrect. Each response was scored for accuracy (0 = incorrect, 1 = correct), with a maximum total possible score of 3 in each task. In the Feedback conditions, if a participant did not answer one item correctly, it was initially scored 0. However, if after feedback they answered correctly, it was scored 1. Therefore, for participants who failed at least one item, we computed both pre and post feedback scores, in order to determine the effectiveness of the manipulation.

*2.2.2: Moral dilemmas*

We used three personal, self, inevitable dilemmas: two selected from the set used by Greene et al. (2004; Crying Baby and Submarine) and one from Moore, Clark, and Kane (2008; Rescue 911)[7].

[7] Moore et al. (2008) distinguished between impersonal/personal, self/other, and inevitable/avoidable dilemmas. A personal dilemma is one in which the person must directly sacrifice another in order to save several lives (e.g. by strangling or throwing from a helicopter) while an impersonal dilemma is less direct (e.g. pushing a button or pulling a lever). In "self" dilemmas, the person who makes the decision will also die if he/she decides not to sacrifice; while "other" dilemmas are those in which

We focused on this kind of dilemmas because Moore and colleagues found that responses to personal and inevitable dilemmas were affected by individual differences in working memory—a cognitive ability that is positively correlated with both CRT (Toplak, West, & Stanovich, 2011) and BNT scores (Cokely et al., 2012). After reading each dilemma, participants indicated whether they thought they should perform the action proposed in each case (e.g., *Do you think that you should asphyxiate your child to save yourself and the other people hidden?*). The response scale ranged from 1 (definitely "no", do not sacrifice), to 6 (definitely "yes", do sacrifice). This scale was used because it allows considering the answer as a dichotomy (1-3, non-utilitarian, 4-6 utilitarian) as well as grading the response (Bartels & Pizarro, 2011). All three dilemmas were translated into Spanish by the first author, who is a native English speaker also proficient in Spanish, and were reviewed by the last author, a native Spanish speaker with excellent knowledge of English.

### *2.2: Procedure*

All materials were implemented as an electronic survey in Unipark (www.unipark.de). Sample A participants were seated at individual computers in the laboratory, whereas Sample B participants completed the study online. In all cases, participants provided informed consent (either in writing or online) before proceeding to the study. Next, they read instructions indicating that they would need paper and pen to work out the problems and write down their answers before entering them on the computer. In the feedback conditions, this allowed participants to check their initial

> the life of the person who makes the decision is not at stake. Finally, dilemmas where the victim will die regardless of the choice are "inevitable", whereas "avoidable" dilemmas are those where the victim will live if the decision is not to sacrifice.

response and change it in their second attempt, if necessary. They were then allocated to one of the four experimental conditions: Sample A: CRT, *n* = 56; BNT, *n* = 48; CRT_Fb, *n* = 90; and BNT_Fb, *n* = 36; Sample B: CRT, *n* = 56; BNT, *n* = 55; CRT_Fb, *n* = 57; and BNT_Fb, *n* = 54 (Fb = feedback groups). Demographics for each of the experimental conditions are provided in Supplementary Materials.

Before starting the experiment, participants completed a practice trial to become familiarized with its structure. The practice trial consisted of one item (either from the CRT or the BNT, depending on the experimental condition), followed by a detailed explanation of the steps required to solve the problem and arrive at the correct answer. This aimed to demonstrate to participants in the feedback conditions that they were not being "tricked", and that there was indeed an error in their answers[8]. Only participants in feedback conditions received feedback if they answered the practice question incorrectly.

After the practice trial, participants were presented either with the three CRT (original or new, for Sample A and B, respectively) or BNT items. In the feedback conditions, participants received feedback if they responded incorrectly, and were given the chance to respond again; this occurred just once for each wrong answer. If despite the feedback they responded incorrectly a second time, the program proceeded to the next question without further comments on their performance (see Figure 1.1, for a flow chart of the experimental procedure). After completing the

[8] In a pilot study we asked participants to comment on any aspect of the experiment they found relevant. Several participants signaled that they knew very well that the feedback was just a "trick" manipulation, and that they were quite confident that their responses were correct. This is congruent with Frederick's (2005) observation that participants with lower CRT scores generally evaluated the problems as easier than those with higher scores.

three items, participants responded to the three moral dilemmas in a randomized order. Finally, participants were asked if they had prior knowledge of any of the CRT or BNT items. Upon completion of the experiment, participants were offered a debriefing session.

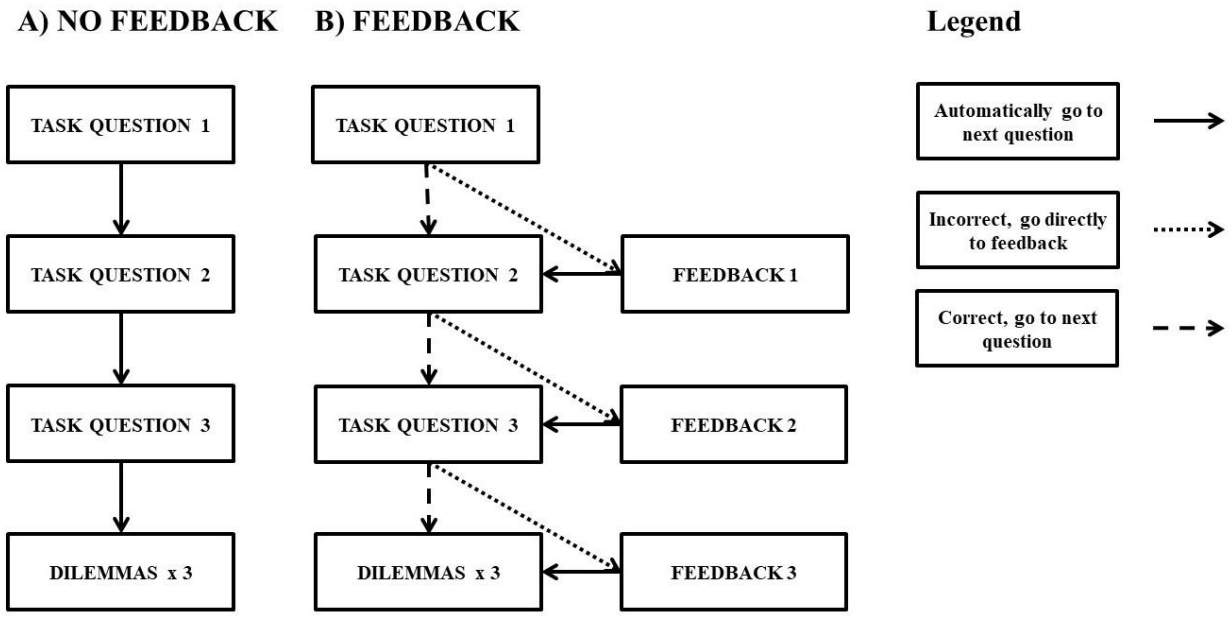


Figure 1.1 Flow chart of the experimental procedure. Participants were allocated to one of the four experimental conditions: CRT, CRT with feedback (CRT_Fb), BNT, or BNT with feedback (BNT_Fb). After responding to each question, participants in conditions without feedback were presented directly with the next question independently of whether their answer was correct. In conditions involving feedback, participants viewed the same question again only if they answered it incorrectly, along with feedback to help them reflect on their answer. Feedback was only given once for each question. After the three task questions, participants read and responded to three moral dilemmas, which were presented in a randomized order.

## 3: Results

### *3.1: Manipulation check: feedback effect on CRT/BNT performance*

As described in the Procedure section, feedback was only provided to participants in the feedback conditions who did not reach a perfect performance in the CRT/BNT task (3 correct responses). These participants entered the present analysis to test whether feedback was effective at improving CRT/BNT performance. After removal of the individuals who did not receive any feedback the analysis sample consisted of 208 participants. This included 121 participants from Sample A (85 female, age $M$ = 19.17, $SD$ = 2.55, range 17 - 36) and 87 from Sample B (53 female, age $M$ = 29.17, $SD$ = 10.01, range 18 – 68).

Figure 1.2 displays the distribution of observations across performance levels (0, 1, 2, or 3 correct responses) before and after the feedback manipulation. Given that people who showed a perfect performance before feedback were excluded from the analysis, there were no observations in level 3 prior to feedback delivery. Results are shown separately for sample A (upper panel) and sample B (bottom panel), and for the CRT and the BNT tasks.

These data were analyzed following a Time (pre/post feedback) x Sample (A, B) x Task (CRT, BNT) design. As shown in Figure 1.2, a consistent distribution displacement to the right – performance improvement– occurred after providing feedback. This was confirmed by a Generalized Linear Mixed-effects model (i.e., Cumulative Link Mixed Model fitted with the Laplace approximation, using the *clmm* function from the *ordinal* R package; Christensen, 2018). Ordinal modeling was used in view of the fact that CRT/BNT performance is not a continuous measure (it ranges from 0 to 3, representing the number of correct responses), and should not be treated as such (Liddell & Kruschke, 2018). A logit link and flexible thresholds were used for analyses. Task (CRT, BNT), Sample (A, B), Time (pre-feedback, post-feedback), and their first- and second-order interactions entered the model as fixed factors, whereas participant was considered a random-effects factor.

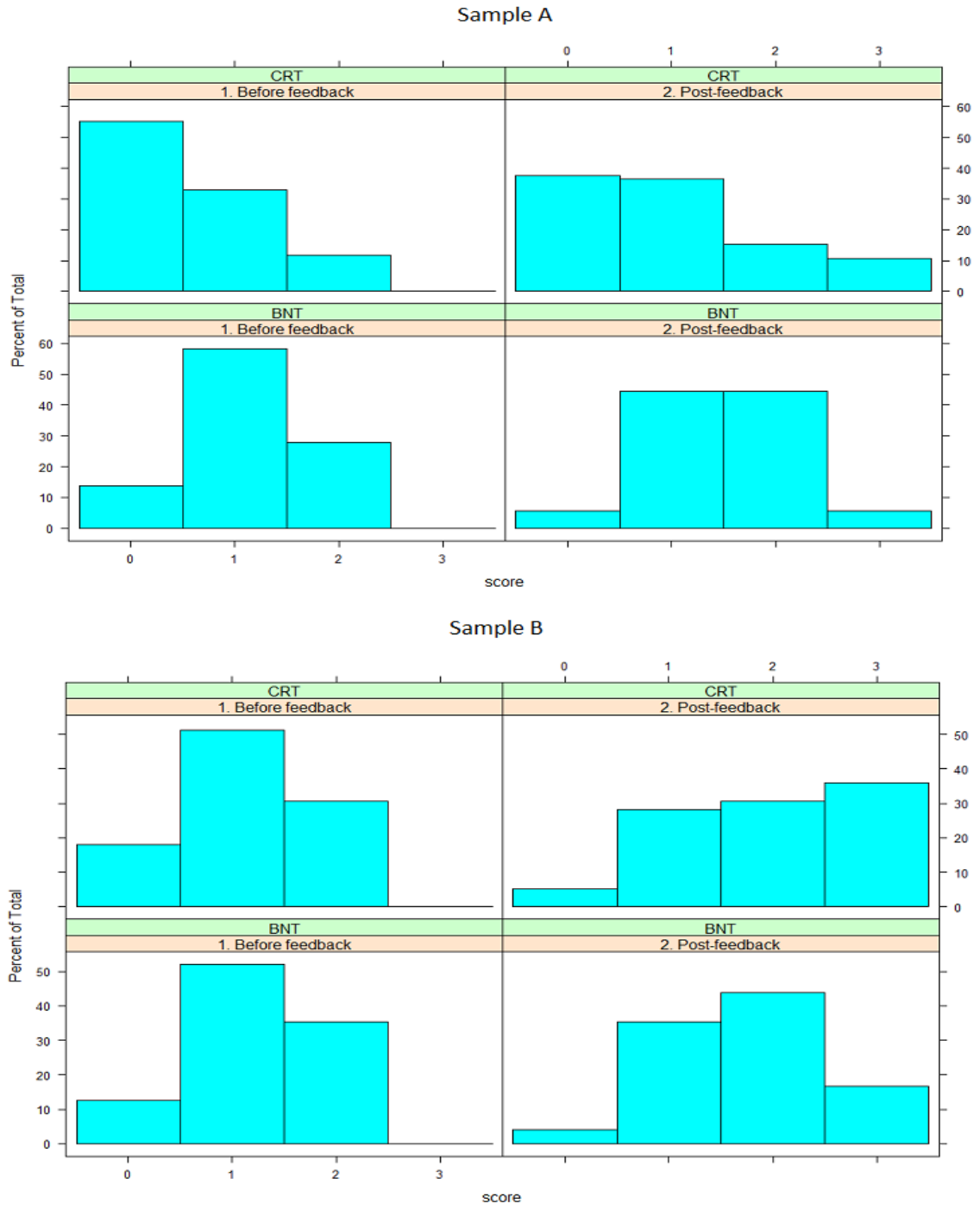


Figure 1.2. Distribution of observations across CRT/BNT performance levels for Sample A and Sample B.

The analysis yielded a significant Time (pre-post feedback) effect (B = 1.612, SE = 0.548, $z$ = 2.944, $p$ = .003, $OR$ = 5.01). Additionally, there was a Task effect (B = -3.745, SE = 0.934, $z$ = -4.009, $p$ < .001, $OR$ = 0.02), with the CRT eliciting on average fewer correct responses than the BNT; and a Task x Sample interaction (B = 3.188, SE = 1.276, $z$ = 2.499, $p$ = .012, $OR$ = 24.23), indicating that the difference between tasks was larger for Sample A than for Sample B participants (or, in other words, Sample B performed the CRT better than Sample A, a difference that was not evident for BNT). Marginal/conditional $R^2$ values for the model were 0.27/0.84.

### *3.2: Matching check: baseline BNT/CRT performance differences across groups*

A supplementary analysis was run to check whether the two Feedback conditions (feedback/no feedback) differed in baseline CRT/BNT performance, i.e., whether CRT/BNT performance differed between the no-feedback group and the feedback group before feedback was delivered. After removal of the individuals who did the task perfectly (n = 9 in Sample A and n = 37 in Sample B), the analysis sample consisted of 406 participants in total. This included 221 participants from Sample A (148 female, age $M$ = 19.70, $SD$ = 2.77, range 17 - 36) and 185 from Sample B (121 female, age $M$ = 27.60, $SD$ = 10.40, range 18 – 68).

With regard to modeling, the analysis paralleled the one described in the previous section, except for the fact that there was no within-subject manipulation and thus no random intercept for participant either. (All predictors were fixed-effects factors, and the *clm* function instead of *clmm* was used for fitting). Task (CRT, BNT), Sample (A, B), and Feedback (feedback, no feedback), and their first- and second-order interactions entered the model as fixed factors. Most importantly, the effect of Feedback condition on baseline CRT/BNT performance (henceforth, *Baseline*) did

not reach significance (B = - 0.307, SE = 0.413, $z$ = - 0.742, $p$ = .458, $OR$ = 0.74). Interactions involving feedback also remained far from significance (min. $p$ = .351). Paralleling the previous analysis, the number of correct responses at baseline was smaller for CRT than for BNT (B = - 1.654, SE = 0.384, $z$ = - 4.306, $p$ < .001, $OR$ = 0.19), and task interacted with sample, again in such a way that Sample B outperformed Sample A in the CRT, but not in the BNT (B = 1.387, SE = 0.558, $z$ = 2.486, $p$ = .013, $OR$ = 4.00, for the interaction). Nagelkerke's $R^2$ for the model was 0.15.

### *3.3: Main analysis: Effects of baseline CRT/BNT performance and feedback on responses to moral dilemmas*

Main analyses were aimed at testing whether receiving feedback for CRT/BNT performance (vs. performing the BNT/CRT task just once) had any effect on participants' responses to moral dilemmas. Again, only participants who did not perform the task perfectly in the first place were included in these analyses. The sample was thus the same as in the previous section.

Figure 1.3 displays the distribution of frequencies across utilitarianism scores (collapsed across dilemmas), for the two samples, the two tasks, and the two feedback conditions. Analyses obeyed to a Sample (A, B) x Feedback (feedback, no feedback) x Task (CRT, BNT) design, with responses to dilemmas as the dependent variable. Generalized Linear Mixed-effects (Ordinal) models were fitted to responses to the three dilemmas. In order to use the three dilemmas as different items for the same construct (stronger/weaker utilitarianism), Dilemma (Crying baby, Submarine, and Rescue 911) was treated as a random-effects factor, along with Participant. Sample (A, B), Feedback (feedback, no feedback), Task (CRT, BNT), and all possible interactions entered

the initial (saturated) model as fixed-effects factors. The score in each dilemma (considered as an ordinal measure) was the output variable.

Given the large number of possible interactions, a backward hierarchical procedure was firstly followed to exclude the interactions that did not contribute to model fit. The removal of the three-way interaction (Sample x Task x Feedback) from the initial (saturated) model improved model fit by a difference in the Akaike Information Criterion (ΔAIC) of - 1.539. Removals of the Feedback x Sample interaction did not hamper model fit either (ΔAIC = - 0.902). However, removal of the Task x Sample (ΔAIC = 3.245) and the Task x Feedback (ΔAIC = 3.615) interactions did reduce model fit and were kept in the final (best-fitting) model. Given that all main factors are involved in the interactions preserved in the final model, none of them can be further removed. The final model thus consisted of Task, Sample, Feedback, Task x Sample, and Task x Feedback in the fixed part.

This model yielded significant effects of Task (B = -1.081, SE = 0.334, $z$ = -3.238, $p$ = .001), Feedback (B = - 0.680, SE = 0.290, $z$ = -2.342, $p$ = .019, $OR$ = 0.51), Task x Feedback (B = 0.877, SE = 0.393, $z$ = 2.230, $p$ = .026, $OR$ = 2.40), and Task x Sample (B = 0.882, SE = 0.398, $z$ = 2.215, $p$ = .027, $OR$ = 2.42). The signs of effect estimates reveal slightly more utilitarian responses, in general, after receiving the CRT, but this effect was mostly attributable to Sample B (who actually did the CRT better than Sample A to begin with). Most importantly, however, participants become more utilitarian after receiving feedback (vs. not receiving it), but this effect was mostly restricted to participants who received feedback for the BNT (Figure 1.3, left panels).

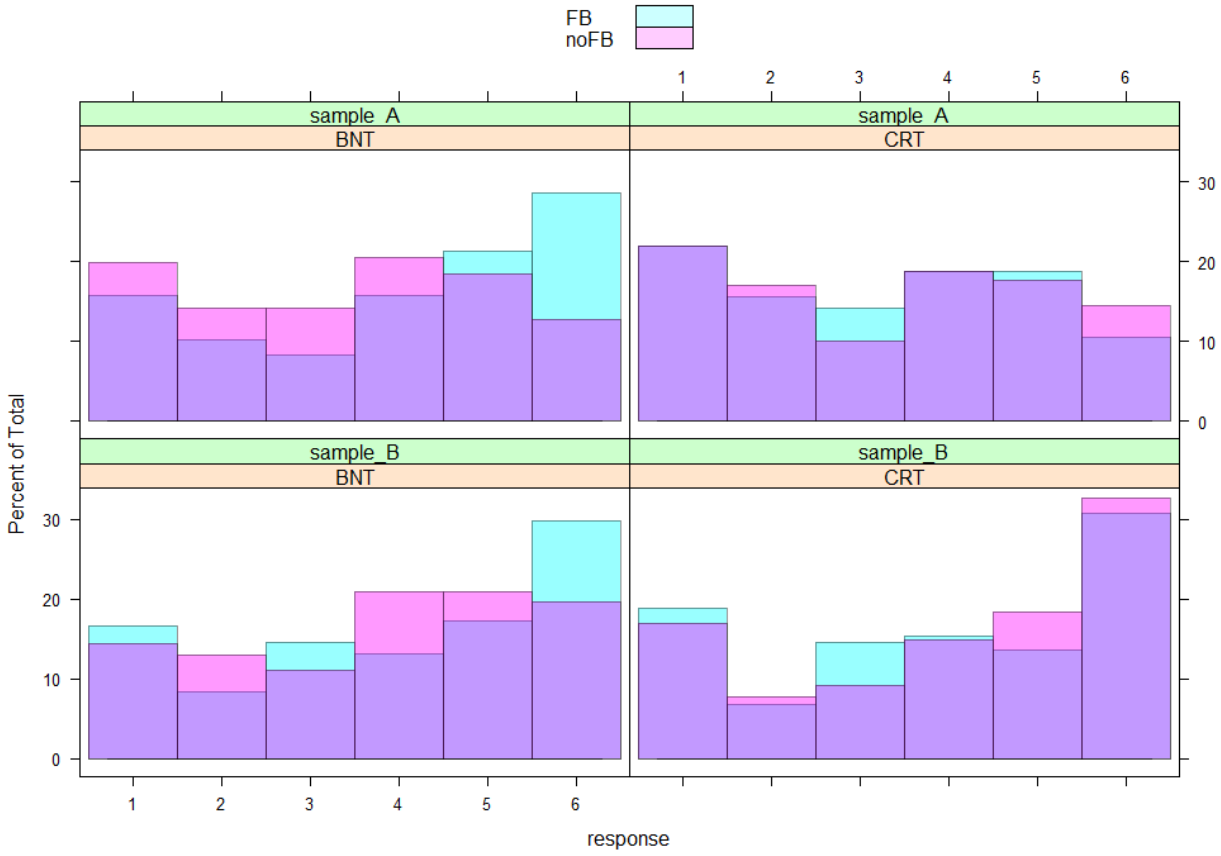


Figure 1.3. Distribution of observations across utilitarianism scores (accumulated across dilemmas) for the two Tasks (BNT/CRT), Samples (A/B) and Feedback conditions (FB/noFB). Note. Purple areas correspond to the overlapping between FB (blue) and noFB (pink) conditions.

As shown in the right panels of Figure 1.3, feedback for the CRT was virtually ineffective to further increase utilitarianism. Marginal/conditional $R^2$ values for the model were 0.04/0.59.

Finally, a supplementary analysis was run to test the well-known relationship between performance in reasoning tasks and utilitarianism. This analysis was performed only in individuals who did not receive feedback for their BNT/CRT performance. Contrasting with previous analyses, participants who performed perfectly at baseline were not removed (so for the present analysis, $n$ = 215). Baseline CRT and BNT scores were separately standardized before entering the model. Responses in the three dilemmas were regressed over initial BNT/CRT performance,

task type (BNT/CRT), and sample (A/B). As in previous analyses, dilemma and participant were considered random-effects factors. Additionally, prior to analysis, BNT and CRT performance were separately standardized to remove scalar differences between them. Utilitarianism scores were higher for sample B individuals (B = 0.845, SE = 0.276, $z$ = 3.069, $p$ = .002, $OR$ = 2.33), and also (and independently) for people with higher initial BNT/CRT scores (B = 0.651, SE = 0.146, $z$ = 4.472, $p$ < .001, $OR$ = 1.92). Utilitarianism scores were not significantly affected by the type of reasoning task used ($p$ = .304). Marginal/conditional $R^2$ values for the model were 0.08/0.61.

**4: Discussion**

The main aim of this work was to shed light on the role of cognitive reflection vs. deliberation on moral decision making. Participants responded to moral dilemmas after completing a task that requires both inhibiting intuitive responses and deliberation (Cognitive Reflection Test, CRT), or a task that recruits deliberative processing, but does not require inhibiting intuitive responses (the Berlin Numeracy Test; BNT). We reasoned that if intuition inhibition (and not just deliberative thinking) increases utilitarian responses, then such responses should be more common among individuals who respond to the CRT compared to the BNT. Additionally, to distinguish between the potential effect of natural reflective traits or thinking dispositions (Baron et al., 2015) and that of induced reflection (Paxton et al., 2012), we manipulated whether participants received feedback on their performance in the CRT or the BNT. Feedback was designed to encourage even those participants who are not naturally reflective to reconsider their responses. We reasoned that, if the inhibition of prepotent intuitive responses drives utilitarian responses to moral dilemmas, then feedback linked to CRT items should lead to more utilitarian responses. In contrast, any effect of BNT on responses to moral dilemmas should not be affected by feedback.

Our data, however, did not fully support these predictions. Although feedback resulted in more utilitarian responses to moral dilemmas, this effect was mostly attributable to feedback on the BNT. The effect was not attributable to differences in baseline task performance. Additionally, both CRT and BNT scores predicted utilitarian responses when feedback was not provided. That performance in the CRT predicts utilitarian decisions is in agreement with a previous study linking cognitive reflection to utilitarian choice (Paxton et al., 2012; but see Sirota et al., 2020 for the absence of effect when using a verbal CRT without numeric component). Our results also show that performance in a numeracy task which has been reported to predict better performance across diverse tasks—such as distinguishing between profound statements and pseudo-profound bullshit (Erlandsson et al., 2018), superior judgment and decision making (Cokely et al., 2012), and meta-cognitive performance (Ghazal et al., 2014)—also predicts utilitarian responding. These results are in agreement with recent evidence that inducing analytical thinking outside the numeric domain, directly by instructing participants to use this thinking mode (Li et al., 2018) or indirectly by inducing processing disfluency (Spears et al., 2018), leads to more utilitarian decisions. More generally, our results are consistent with the notion that deliberative reasoning predicts utilitarian decisions and provide support for the deliberative-utilitarian link (vs. a reflective-utilitarian link). Specifically, our data suggest that an increase in utilitarian responses may be induced by performing a numeric task that triggers deliberative thinking. However, generalizing our results to any cognitive (i.e., non-numeric) task may be premature.

Concerning the effect of feedback, participants in feedback conditions likely engaged in 'deeper deliberation' to reach an accurate response (Ghazal et al., 2014), leading to more utilitarian moral decisions. The fact that feedback mostly affected responses to moral dilemmas when it was linked to the BNT (and not to the CRT) casts more doubts upon the reflective-utilitarian link.

However, we cannot rule out the possibility that responses to moral dilemmas were affected to some extent by the numerical nature of both tasks (see Sirota et al. 2018). The requirement to complete either task before responding to dilemmas could potentially make the numeric aspects in dilemmas more salient, thus increasing utilitarian decisions (Bialek & Terbeck, 2016; Bialek, & De Neys, 2017). Feedback may further increase attention to the numerical nature of the tasks. This account, however, would not explain why the effect of feedback on responses to moral dilemmas was larger in the case of BNT, even though feedback improved performance for both tasks. That feedback did not increase utilitarian responses among participants who completed the CRT could be related to the effects of the lures on this task. In order to respond correctly to CRT items, participants are first drawn to the intuitive response (induced by the lures) before reflecting, and correcting the error. We may speculate that such lures strongly attract people's attention to intuitive incorrect answers, thus limiting the process of deliberation and, consequently, its impact on moral judgments.

It is also worth considering other criticisms of the deliberative-utilitarian connection which are more philosophical and directed toward the interpretation of the dilemmas themselves. Although we have concluded that correct answers on the CRT and BNT tasks predict utilitarian decisions, some might argue that these tasks only predict a particular, consequentialist aspect of utilitarianism. Specifically, Kahane et al., (2018) have argued that utilitarianism advocates both doing the least harm but also bringing about the greater good including altruistic acts – an aspect we cannot measure using the specific (sacrificial) dilemmas we chose. As noted in the Introduction, the utilitarian response might be chosen, for instance, by a participant exhibiting the callous affect psychopathic personality trait. However, Conway et al. (2018) found that people who make more utilitarian decisions on sacrificial moral dilemmas also reported being more

utilitarian in other aspects of utilitarianism, such as giving more money to charity and being more pro-social. Another study also found that participants who score higher on the BNT and CRT reported giving more to charity and/or volunteering in the past year (Erlandson et al., 2018). In summary, people who are more deliberative may be more utilitarian not only in the instrumental harm aspects, but also in the altruistic aspects of utilitarianism.

### *4.1: Limitations*

One limitation of our work is that we did not measure participants' mood, which may affect responses to moral dilemmas. It is possible that the happiness resulting from the knowledge that a question was answered correctly (for example, by not receiving negative feedback) increased utilitarian responses to moral dilemmas to some extent (Gawronski et al., 2018). It is also possible that the difficulty of the tasks and/or the negative feedback gave rise to negative mood, possibly affecting participants' responses to moral dilemmas to some extent (Pastötter et al., 2013).

Another limitation of our work is that it focused on sacrificial dilemmas, which are often considered to be unrealistic (e.g. Bauman et al., 2014) or ill-suited to capture different dimensions of utilitarianism (Kahane, et al, 2018). Yet, some of these dilemmas resemble real life situations. For instance, programming an autonomous or driver-less car involves decisions concerning whether to save a set of pedestrians crossing the street or the passenger of the car, in the event of an unavoidable accident (Bonnefon et al., 2016; Powell et al., 2016). Future research could examine whether the pattern of results documented here generalizes to a more diverse set of ecological dilemmas.

Finally, future research could also examine whether moral judgments and decisions are affected by feedback provided in connection to other numerical tasks without lures or even to

lexical tasks, such as a syllogistic one (Baron et al., 2015). This could contribute to determine to what extent deliberation per se vs. performing tasks involving a numeric aspect associates with utilitarian decisions. This could also help to understand to what extent moral judgments and decisions are affected by performance feedback for different tasks.

### *4.2 Conclusions*

The current findings suggest that individual differences on both CRT and BNT performance predict utilitarian responses to moral dilemmas. While feedback enhanced performance in both tasks, it only increased utilitarian responses in connection to the BNT. These results suggest that performance in a numeric task requiring deliberative thinking (CRT or BNT) may predict utilitarian responses to moral dilemmas, and that the inhibition of intuitive wrong responses (reflection) may not always be necessary to induce utilitarian decisions.

# Study 2:

# Disfluent fonts lead to more utilitarian decisions in moral dilemmas

Publication:

Spears, D., Fernández-Linsenbarth, I., Okan, Y., Ruz, M., & González, F. (2018). Disfluent fonts lead to more utilitarian decisions in moral dilemmas. *Psicológica, 39(1), 41–63*. https://doi.org/10.2478/psicolj-2018-0003

## 1: Introduction

People are often faced with difficult moral dilemmas that can have important life changing consequences. Accordingly, an increasing body of work has sought to identify the different factors that affect moral judgments and decisions, as well as the processes underlying people's moral choices. The dual process theory of moral decision-making (Greene et al., 2001; Greene et al., 2004) is a prominent psychological approach that distinguishes between two different systems (e.g., Kahneman & Frederick, 2002; Sloman, 1996), namely an intuitive emotional system, which tends to be fast and automatic, and a controlled cognitive one, which tends to be slow and deliberate. According to this theory, decisions involving intuition and emotion generally support *deontological* judgments in moral dilemmas, favoring the essential rights of a person (e.g., one should not sacrifice an individual, even if this implies saving a larger number of people). Conversely, controlled cognitive processing favors *utilitarian* judgments, which lead to the greater good (e.g., one should aim to save the larger number of people, even if this means sacrificing an individual)[9].

Numerous studies manipulating cognitive and emotional factors have provided support for the claim that utilitarian and deontological responses to moral dilemmas rely on dissociable cognitive processes. To illustrate, the prevalence of utilitarian responses has been found to increase as a result of manipulations that promote analytic over intuitive emotional processes. These include allowing participants more time to respond (Suter & Hertwig, 2011), presenting dilemmas in a

[9] While Greene et al.'s (2001, 2004) dual process model has provided the basis for a wide range of interesting studies about the role that emotion and cognition play in moral decision-making, it is not free from criticism. Many critiques focus on the vision of emotion and cognition working as two dissociable, antagonist and competitive cerebral processes, and they propose instead an interaction of both in moral decision-making (e.g., Duke & Bègue, 2015; Gürçay & Baron, 2017; Moore, Clark, & Kane, 2008).

foreign language (Costa et al., 2014, 2017), or requiring participants to complete the *Cognitive Reflection Test* (CRT; Frederick, 2005) before they respond to dilemmas (Paxton et al, 2012). Utilitarian responses are also more common among individuals with a rational style of thinking (Bartels, 2008) and with higher working memory capacity (Moore et al., 2008). Conversely, manipulations that hamper controlled cognitive processes such as induced stress (Youssef et al., 2012) and cognitive load (Trémolière et al., 2012) often result in a higher prevalence of deontological responses. Taken together, these studies suggest that increased analytic thinking constitutes a key path to utilitarian responses to moral dilemmas (but see Baron et al., 2015;Gürçay & Baron, 2017).

Processing fluency (i.e., the experienced ease with which one processes information) may be another factor influencing moral reasoning, as it is a pervasive metacognitive cue that can have a substantial influence on reasoning and judgment across many domains (Alter & Oppenheimer, 2009). These include, for example, judgments of truth (Reber & Schwarz, 1999) and confidence (Alter et al., 2007), where fluent stimuli are judged as truer and inspire more confidence than the same stimuli presented in a disfluent way. Studies have used a vast array of fluency manipulation techniques, including the visual ease with which stimuli are perceived (Alter et al., 2007; Reber & Schwarz, 1999), the ease with which stimuli are retrieved from memory (Bornstein & D'Agostino, 1992) and priming procedures (Swami et al., 2014). A common processing fluency manipulation has been to present stimuli either in an easy- or difficult-to-read font (Alter & Oppenheimer, 2009). Fonts that are harder to read are thought to increase experiences of processing difficulties (i.e., disfluency), which in turn can serve as a cue that more deliberative and effortful cognitive processes need to be activated (Alter et al., 2007; Alter & Oppenheimer, 2009). Supporting this idea, there is evidence that processing disfluency can lead to better performance on different

cognitive tasks. For example, presenting information in a difficult-to-read font improves performance on the CRT and syllogistic reasoning tasks (Alter et al., 2007), the Moses Illusion task (Song & Schwarz, 2008; Swami et al., 2014) and beyond the laboratory, in educational settings (Diemand-Yauman et al., 2011). Processing disfluency is also associated with lower belief in conspiracy theories (Swami et al., 2014). Taken together, the findings reviewed suggest that processing disfluency can trigger *analytic thinking* strategies, which may help overcome initial intuitive responses[10].

However, it is currently unclear whether perceptual disfluency can also affect people's responses to moral dilemmas. To the best of our knowledge, only one study (Laham et al., 2009) has examined a related issue, namely the role of fluency on evaluations of moral transgressions conducted by others (e.g., a family eating their dead dog for dinner). In this study, transgressions presented in an easy-to-read font were judged as less morally wrong than those presented in a difficult-to-read font. The authors argued that this tendency was due to the fact that the experience of fluency is hedonically marked, which would trigger a positive affective state. Relatedly, there is evidence that processing fluency can affect interpersonal evaluations, leading to more positive evaluations of individuals who are processed fluently, and more negative evaluations of those who are processed disfluently (Lick & Johnson, 2015). Taken together, these studies suggest that fluency may indeed affect some aspects of moral reasoning. However, it is unclear whether fluency can affect how people respond to dilemmas involving a proposed utilitarian action (for a distinction

[10] Although several studies have found that disfluent fonts improve performance on different cognitive tasks, it should be noted that some studies have failed to replicate this effect (Meyer et al., 2015), or found it limited to individuals with high cognitive ability (Thompson et al., 2013).

between deciding in moral dilemmas and reacting to moral transgressions, see Monin et al., 2007). Our main goal in the current work was to investigate this issue.

As noted earlier, analytic thinking is thought to prompt utilitarian responses to moral dilemmas. If disfluency enhances analytic, controlled cognitive processing when people are faced with such dilemmas, then disfluency might result in more utilitarian responses. To investigate whether this is the case or not, we presented participants with a series of moral dilemmas, and manipulated the font in which these were displayed. Based on the work reviewed above, we hypothesized that participants would rate the proposed actions as more appropriate (i.e., more utilitarian responses) when dilemmas were displayed using a difficult-to-read (disfluent) font than when they were displayed using an easy-to-read (fluent) font (H1).

Additionally, we aimed to examine whether a potential effect of disfluency would generalize across different dilemma types, or whether it would be restricted to certain types of dilemmas. Based on Greene et al. (2001, 2004), we distinguished between personal dilemmas (where the proposed action leads to serious bodily harm to the victim or victims, and this harm is not the result of deflecting an existing threat) and impersonal ones (involving the deflection of a threat but no agency). As compared to impersonal dilemmas, personal ones often elicit heightened emotional reactions leading to more frequent deontological responses. Thus, manipulations promoting analytic processing can have a stronger effect on personal dilemmas, as such processing may help to overcome the initial emotional response triggered by these scenarios (Costa et al., 2014; Duke & Bègue, 2015; Koenigs et al., 2007). Therefore, we further hypothesized that any effect of processing fluency would be stronger for personal than for impersonal dilemmas (H2).

To exclude potential confounds of any effects of fluency, we recorded the time that participants spent reading the dilemmas, as well as participants' self-reported mood. Utilitarian

responses can be more common when people are given more time to respond (Suter & Hertwig, 2011; but see Gürçay & Baron, 2017). Thus, dilemmas written in a disfluent font may elicit more utilitarian responses than those written in a fluent font simply because reading the former type of font takes longer, and not because of the metacognitive cue provided. Additionally, we assessed participants' mood because disfluent fonts may lead to a more negative mood (Alter et al., 2007). Negative moods can prompt more systematic processing (Schwarz, Bless, & Bohner, 1991). Thus, disfluency could also result in more utilitarian responses through its negative effect on mood, independently of the metacognitive experience of fluency itself. Indeed, as noted earlier, there is some evidence that fluency may elicit positive affect, which might translate into more positive evaluations of moral transgressions or of other people (Laham et al., 2009; Lick & Johnson, 2015). There is also evidence that positive affect can lead to more utilitarian responses to moral dilemmas (Valdesolo & DeSteno, 2006). Hence, we examined whether our fluency manipulation induced any changes in mood by registering participants' self-reported mood after both fluent and disfluent blocks of moral dilemmas.

## 2: EXPERIMENT 1

### *2.1: METHODS*

#### *2.1.1. Participants.*

Fifty-one undergraduate students (38 female, age range 17-34, $M$ = 18.76, $SD$ = 2.33) recruited from the Faculty of Psychology of the University of Granada took part in the study in exchange of course credit. All the procedures were approved by the local Ethics Committee on Human Research.

*2.1.2. Materials and Apparatus*.

All materials were presented using the survey programming software Unipark (www.unipark.de).

*2.1.3. Dilemmas*[11]

Two short dilemmas without moral content were used as practice trials, to familiarize participants with the structure of the decision-making task and with the fluent and disfluent conditions. The dilemmas (*Brownies* and *Standard Turnips*) were taken from Greene et al. (2004). They were followed by a question in which participants had to rate the appropriateness of the proposed action using a scale ranging from 1 (*definitely no*) to 6 (*definitely yes*).

We used four dilemmas, including two fillers (*Stock Tip* and *Taxes*), and two moral (*Crying Baby* and *Burning Building*) dilemmas, taken from Moore et al. (2008). The two moral dilemmas had an almost identical number of words (*Crying* Baby = 126; *Burning* Building = 129), and each of these dilemmas had both a personal and an impersonal version (with exactly the same number of words), resulting in a total of six dilemmas. In all cases participants were asked to indicate how appropriate they felt the proposed action was (e.g., *Would it be appropriate for you to asphyxiate your child in order to save yourself and the other hidden people?*) using the same response scale as in the practice dilemmas[12].

[11] All dilemmas used in this study are included in Appendix A.2.

[12] The moral dilemmas were sacrificial and self-benefit (the life of the person who commits sacrifice is also in danger). They were slightly modified to meet the criteria of *unavoidable death* (the potential victim of the sacrifice would die even if no action were taken). Moore et al. (2008) found that this kind of dilemmas was more sensitive to capture differences in participants' working memory capacity and therefore they may be more sensitive to any effect of manipulating fluency.

All practice trials, dilemmas, and associated questions were written using Myriad Web 10-point font. In the disfluent font condition, the words were at a 15% dark scale and italicized (see Alter et al., 2007, for a similar procedure).

This is an example of the fluent font.

*This is an example of the disfluent font.*

*Mood questions*

Following Alter et al. (2007), participants were asked to evaluate their mood on a response scale ranging from 1 (*very sad*) to 7 (*very happy*). Mood measurements were taken at three different points in the experiment, as will be described below.

**2.2. Design and Procedure.**

We used a 2 (Fluency: Fluent vs. Disfluent) x 2 (Moral Dilemma Type: Personal vs. Impersonal) repeated measures design. Participants saw all six dilemmas, presented in two blocks, each consisting of three dilemmas: the personal version of one of the moral dilemmas, the impersonal version of the other moral dilemma, and one of the fillers dilemmas. That is, participants were presented with one block composed of the personal version of the *Crying Baby*, the impersonal version of the *Burning Building*, and the *Stocktip* dilemma, and another block composed of the impersonal version of the *Crying Baby*, the personal version of the *Burning Building*, and the *Taxes* dilemma. The order of presentation of the dilemmas within each block was randomized. Each participant saw one block written in the fluent font and the other block written in the disfluent font. The order of blocks and fluency conditions was counterbalanced. To examine whether the specific order in which participants saw the two blocks of dilemmas interacted with any effects of fluency, we included order as a factor in the analyses reported below.

Upon arrival, participants sat in front of individual computer screens. They read and signed a consent form and subsequently read instructions asking them to imagine that the described situations were real, and that no other options (besides the two presented) were possible. To ensure that participants read and considered each dilemma, they were informed that they would have to answer questions about the content of the dilemmas, after all of them had been presented (e.g., *'In the story of the baby, where were the people seeking refuge found?'* – the basement).

Before starting the experimental blocks, participants answered basic demographic questions and responded to the first mood question. Next, they were given the practice trials, one written in the fluent font (*Brownies*) and the other one in the disfluent font (*Standard Turnips*), and rated the appropriateness of the proposed actions using the response scale described above. Participants were then presented with the first block of dilemmas. In all cases they were asked to read the dilemma and click on a 'continue' button to proceed to another screen displaying the corresponding question and response scale. Times spent reading each dilemma were unlimited and were recorded.

Next, participants rated their mood again and were subsequently presented with the second block of dilemmas, following the same procedure used with the first block. The font used and the identity of dilemmas varied according to the counterbalancing of both factors.

Finally, participants were asked to respond to the content questions concerning the content of dilemmas (two for the fillers and two for the moral dilemmas), followed by the final mood assessment.

**3: RESULTS**

First, to test our hypotheses concerning the effect of fluency we ran a 2 (Fluency: fluent vs. disfluent, within-subjects) x 2 (Dilemma Type: personal vs. impersonal, within-subjects) x 2 (*Order of experimental blocks, between-subjects, fluent first, $n_1$ = 25, disfluent first, $n_2$ = 26*) mixed Analysis of Variance (ANOVA) on participants' responses to the dilemmas (see Table 2.1 for the means and standard errors for each combination of Fluency, Dilemma Type, and Order). This analysis yielded a main effect of Fluency, $F$ (1, 49) = 5.60, $p$ = .02, $\eta_p^2$ = .10, reflecting that the disfluent condition was associated with higher acceptability ratings than the fluent condition (Disfluent $M$ = 3.29, $SE$ = .15; Fluent $M$ = 3.01, $SE$ = .18), supporting hypothesis H1. However, this pattern was qualified by an interaction between Fluency and Order, $F$ (1, 49) = 5.60, $p$ = .02, $\eta_p^2$ = .10. A post hoc Tukey's test for unequal group sizes revealed that the effect of fluency was significant when the disfluent block was presented first, $p$ = .008, but not when the fluent block was presented first, $p$ = 1.00 (see Figure 2.1). The analysis also revealed a main effect of dilemma type, $F$ (1, 49) = 50.18, $p$ < .001, $\eta_p^2$ = .51, suggesting that participants rated the actions proposed in impersonal dilemmas as more acceptable than those proposed in personal ones, as expected (Impersonal $M$ = 3.59, $SE$ = .16; Personal $M$ = 2.71, $SE$ = .17). However, we did not find evidence for our hypothesis that effects of processing fluency would be stronger for personal dilemmas than for impersonal ones (H2), as the interaction between Fluency and Dilemma type was not significant, $F$ < 1. No other main effect or interactions were significant, highest $F$ (1, 49) = 1.48, $p$ = .23, $\eta_p^2$ = .03.

**Table 2.1. Responses to moral dilemmas.**

| Dilemma | Order | *M (SE)* |
|---|---|---|
| *Fluent* | | |
| Personal | Fluent- Disfluent | 2.72 (.31) |
| | Disfluent_Fluent | 2.54 (.30) |
| Impersonal | Fluent- Disfluent | 3.20 (.31) |
| | Disfluent_Fluent | 3.58 (.30) |
| *Disfluent* | | |
| Personal | Fluent- Disfluent | 2.32 (.30) |
| | Disfluent_Fluent | 3.27 (.29) |
| Impersonal | Fluent- Disfluent | 3.60 (.31) |
| | Disfluent_Fluent | 3.96 (.30) |

Note. Acceptability ratings means (M) and standard errors (SE) as a function of Fluency (Fluent vs. Disfluent), Type of Dilemma (Personal vs. Impersonal), and Order of blocks (Fluent or Disfluent first).

Although the analyses above supported our hypothesis concerning the effect of fluency (H1), from these analyses alone we cannot firmly conclude that disfluent fonts lead to more utilitarian responses. As noted above, differences between fluent and disfluent dilemmas were only observed when the disfluent dilemmas were presented first. Thus, it is possible that this effect is due to the order of presentation and not to the fluency manipulation itself. In order to solve the potential confound between our fluency manipulation and order, we selected only the first block for each

participant (that was either fluent or disfluent) and analyzed fluency as a between subject variable[13] (means for Personal dilemmas, Fluent *M* = 2.72, *SE* = .30, Disfluent *M* = 3.27, *SE* = .27; Impersonal, Fluent *M* = 3.20, *SE* = .31, Disfluent *M* = 3.96, *SE* = .30). The 2 (Dilemma Type: personal vs. impersonal, within-subjects) x 2 (Fluency: fluent vs. disfluent, between-subjects) mixed ANOVA yielded a significant main effect of Dilemma Type, $F$ (1, 49) = 4.62, $p$ = .04, $\eta_p^2$ = .09, indicating that utilitarian responses were higher for impersonal dilemmas (Impersonal *M* = 3.58, *SE* = .21; Personal *M* = 2.99, *SE* = .21). This analysis also revealed a marginally significant main effect of fluency, $F$ (1, 49) = 4.00, $p$ = .05, $\eta_p^2$ = .07, reflecting higher acceptability ratings for disfluent dilemmas than for fluent ones, as expected (Disfluent *M* = 3.61, *SE* = .23; Fluent *M* = 2.96, *SE* = .23). The interaction between dilemma type and fluency, however, was not significant, $F < 1$. Thus, we basically obtained the same pattern of results in this latter analysis without the potential confound arising between fluency and order of presentation.

Next, we examined whether the time that participants spent reading dilemmas varied as a function of fluency or dilemma type. To this end we ran another 2 (Fluency: fluent vs. disfluent, within-subjects) x 2 (Dilemma Type: personal vs. impersonal, within-subjects) x 2 (Order of experimental blocks, between-subjects) mixed ANOVA on reading times. As reading-time frequency distributions departed from normality, we log-transformed the data ($\log_{10}$), which improved both skewness and kurtosis. The ANOVA on log-transformed times revealed a main effect of dilemma type, $F$ (1, 49) = 10.34, $p$ = .002, $\eta_p^2$ = .17, whereby participants took longer to read impersonal dilemmas than personal ones (Impersonal *M* = 4.51, *SE* = .02; Personal *M* = 4.44,

13

*SE* = .03). However, the main effect of fluency was not significant, $F$ (1, 49) = 1.47, $p$ = .23, $\eta_p^2$ = .03.

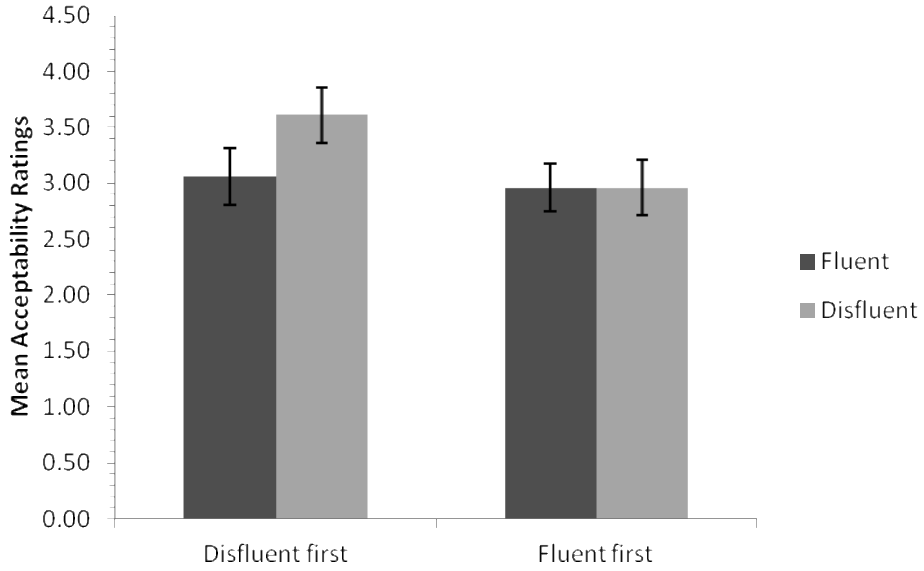


Figure 2.1. Mean acceptability ratings for the moral dilemmas presented in fluent and disfluent conditions, as a function of the order of experimental blocks (disfluent vs. fluent first). Error bars represent ± 1 SEM (standard error of the mean).

The analysis yielded an interaction between Fluency and Order that approached conventional levels of significance¸ $F$ (1, 49) = 3.92, $p$ = .053, $\eta_p^2$ = .07. However, post hoc tests did not reveal any significant difference between means, lowest $p$ = 0.12.

Interestingly, this analysis also revealed an interaction between Fluency and Dilemma type that approached conventional levels of significance, $F$ (1, 49) = 3.85, $p$ = .055, $\eta_p^2$ = .07 (see Figure 2.2). A Tukey's HSD post hoc test revealed a significant difference in reading times between personal fluent and impersonal fluent ($p$ = .01), as well as between personal fluent and impersonal disfluent dilemmas ($p$ = .03). No other differences were significant ($p$s > .08). Therefore, we did not find evidence that participants spent longer overall reading the disfluent dilemmas than the

fluent ones, either for personal or for impersonal dilemmas. The finding that participants took less time to read personal vs. impersonal fluent dilemmas is consistent with the notion that personal dilemmas produce quicker emotional responses that trigger deontological responses.

A final analysis was performed to examine whether Fluency had an effect on participants' self-reported mood. To this end, we ran a mixed ANOVA with Timing of assessment (3 levels: Before, Middle, and End) as within-subject factor, and Order (Fluent first vs. Disfluent first) as the between subjects factor. This analysis revealed a main effect of Timing of assessment, $F$ (2, 98) = 31.19, $p < .001$, $\eta_p^2 = .39$ (Before $M = 5.22$, $SE = .12$; Middle $M = 4.45$, $SE = .15$; End $M = 4.53$, $SE = .13$). A Tukey's HSD post hoc test revealed that participants' mood was more positive before starting the experiment ($M = 5.22$),

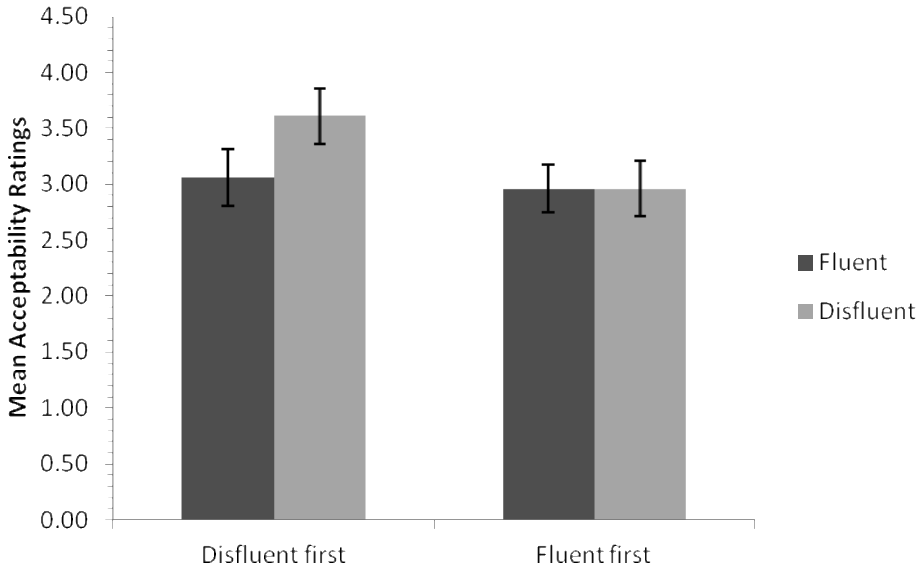


Figure 2.2 Mean log-transformed reading times for personal and impersonal moral dilemmas, as a function of the fluency manipulation. Error bars represent ± 1 SEM (standard error of the mean).

than both after the first block of dilemmas ($M = 4.45$), $p < .001$, and at the end ($M = 4.53$), $p < .001$; there was no significant change between those last two ratings, $p = .74$. The main effect of Order and the interaction between Order and Timing of assessment were not significant, largest $F$

< 1. That is, participants' mood became more negative after the first block and stayed that way after the second block, independently of whether fluent or disfluent fonts were presented first. This suggests that disfluency itself did not lead to a more negative mood (or fluency to a more positive one). A further analysis contrasting pre and post mood measures (before and after the first block of dilemmas, within-subject) being this fluent or disfluent (between-subject) confirmed a reduction of mood from pre to post measures, $F$ (1, 49) = 54.51, $p < .001$, $\eta_p^2 = .53$, but not a difference between fluent and disfluent blocks, $F < 1$, or a block by mood interaction, $F < 1$. Therefore, we did not find evidence of a mood change induced by fluency or disfluency, just a slight reduction in mood from the start to the end of the completion of the first block of trials (Before fluency $M$ = 5.16, $SE$= .17; After fluency $M$ = 4.32, $SE$ = .21; Before disfluency $M$ = 5.27, $SE$ = .16; After disfluency $M$ = 4.58, $SE$ = .20).

## 4: DISCUSSION

The goal of this experiment was to study the effect of disfluent fonts on judgments about moral dilemmas. We hypothesized that when participants read dilemmas in a difficult-to-read (disfluent) font vs. an easy-to-read (fluent) font, they would make more utilitarian decisions. We further hypothesized that this effect would be stronger for personal dilemmas than for impersonal ones. Our results supported our first hypothesis but not the second one. That is, participants endorsed more utilitarian actions in the disfluent condition than in the fluent condition across dilemma types, although they did so only when the disfluent block of dilemmas was presented in first place. However, analyses focusing on the first block of dilemmas revealed a similar pattern (i.e., more utilitarian responses in the disfluent condition), suggesting that the observed effect was

not merely an artifact of the order of presentation of dilemmas. To the best of our knowledge, this is the first reported effect of processing fluency on decision-making concerning moral dilemmas.

Taken together, these findings support and extend previous work pointing to processing fluency as a metacognitive cue that can have a substantial influence on reasoning and judgments across many domains. Our findings are consistent with previous work demonstrating that disfluency enhances *analytic thinking*, thus improving performance in cognitive tasks (Alter et al., 2007; Diemand-Yauman et al., 2010; Song & Schwarz, 2008). Increased analytic thinking can also promote utilitarian responses to moral dilemmas (Bartels, 2008; Moore et al., 2008; Paxton et al., 2012). In our study, dilemmas written in disfluent fonts may have prompted analytic thinking in participants, leading to more utilitarian decisions.

Another possible explanation of our findings relates to the *perceived psychological distance* between stimuli, and how abstractly people represent them. Indeed, there is evidence that processing disfluency can increase perceived distance and lead to more abstract representations (Alter & Oppenheimer, 2008). At the same time, individuals in more abstract mindsets who perceive some distance between themselves and the actions proposed in moral dilemmas may be more likely to give utilitarian responses, as they might focus more on the desirable consequences of the action than on the action itself. Supporting this notion, Aguilar et al. (2013) found that participants who were told that the proposed action (performing surgery on a man and sacrificing his life to save the lives of thousands) would happen in the distant future made more utilitarian choices than those who were told that it would happen in the near future.

In a similar vein, it could be argued that disfluent fonts prompted participants to be in more abstract mindsets in our study, resulting in more utilitarian responses. However, if this were the case, effects of disfluency should be stronger for personal dilemmas than for impersonal ones, as

the killing proposed can be considered as more direct (Moore et al., 2008). Responses to personal dilemmas should thus be more affected by manipulations that increase the psychological distance between the action and its effect, making the killing appear more abstract and less direct. Our results, however, did not support this expectation, as disfluency increased utilitarian responses similarly for both personal and impersonal dilemmas.

Another potential alternative explanation to our finding that the disfluent font was associated with more utilitarian responses is that participants had to slow down to read such font. Indeed, manipulations that encourage people to spend more time reading moral dilemmas can increase the prevalence of utilitarian responses (Suter & Hertwig, 2011). However, our results also revealed that fluency itself did not reliably affect the time that participants spent reading dilemmas or their reported mood. This implies that the observed effect of disfluency cannot be attributed to the potential intervention of these potentially confounding factors. Interestingly, the results showed that, although there was no effect of fluency on reading time for either the impersonal or the personal dilemmas, participants took less time to read the fluent personal dilemmas than the fluent impersonal ones. This difference disappeared for the disfluent dilemmas where both personal and impersonal dilemmas recruited similar amounts of time to be read. This is consistent with the hypothesis that disfluency may trigger analytic processing in personal dilemmas, which is necessary to overcome the initial automatic emotional response of harm aversion, increasing processing time.

Similarly, our findings suggest that the observed effect of fluency on moral responses was not driven by a mood change induced by our fluency manipulation. In our study, mood was slightly more negative after the completion of the first block of trials, regardless of whether this block was fluent or disfluent. This suggests that our one-item measure was sensitive enough to detect this

reduction following the first block of dilemmas, with independence of the font used. However, we did not find support for the notion that the experience of fluency triggers a positive affective state (see Laham et al., 2009; Lick & Johnson, 2015) or that disfluency triggers negative affect (Alter et al., 2009). One reason for these discrepancies with previous work may lay on the nature of our moral scenarios, which present difficult situations involving emotional reactions which may counteract the effect of any affect induced by fluency manipulations.

As with all studies, our investigation has some limitations that could be addressed in future work. First, it would be necessary to use additional disfluency manipulations beside perceptual fluency and investigate whether the results converge with these. Second, the dilemmas used may not reflect situations that people will generally encounter in real life, and thus it is not possible to determine to what extent our findings would generalize to more realistic situations. However, current technological developments such as self-driving cars have created real decision situations that resemble to some extent the kind of dilemmas used here (e.g., Bonnefon et al., 2016). Furthermore, there is evidence that the patterns of responses found for hypothetical dilemmas involving death generalize to dilemmas involving more common kinds of harm such as emotional and economic harm (Gold et al., 2013). Third, in the current study we used only four moral dilemmas. While the use of a limited number of dilemmas is not an unusual practice in the literature of moral decision making (e.g., Costa et al., 2014; Duke & Bègue, 2015), future research should include more diverse dilemmas to assess the generalizability of our findings. Finally, in the current work we did not directly assess the mechanisms underlying the effect of disfluency on utilitarian responses. Future work should examine this issue more directly. This could also help to shed light on our finding that the effect of disfluency held only when disfluent fonts were presented first.

The present work adds evidence to previous studies that highlight processing fluency as a pervasive metacognitive cue that affects a wide range of human judgments. Our findings suggest that increasing processing disfluency by manipulating the font used to present information not only reduces cognitive bias and promotes analytic thinking in mathematical or logic tasks, but can also affect judgments and decisions in the moral domain, biasing people's responses towards utilitarian decisions.

# Study 3:

# Do cognitive control tasks increase utilitarian decisions on moral dilemmas?

## 1. INTRODUCTION

The use of sacrificial moral dilemmas, where a person must either kill one person to save five other people or else do nothing thus letting those five people die, has significantly grown in scientific research to study moral cognition. Early research using these dilemmas claimed that the areas of the brain that activated when a person made the decision to sacrifice another (utilitarian choice) are similar to those activated when participants perform tasks that manipulate cognitive control. An example of this are incongruent trials on a Stroop task (Stroop, 1935), where people must name the ink color of a word that denotes a different hue (Greene et al., 2001; Greene et al., 2004). The commonality of brain activations for responses to moral dilemmas and cognitive control tasks in regions such as the inferior parietal lobe, Anterior Cingulate Cortex, and dorsolateral Prefrontal Cortex, has been used as an argument to claim that higher lever, Type 2 cognition, underlies utilitarian responses made in these situations (Greene et al., 2004; Li et al., 2019; Patil et al., 2021; Paxton et al., 2012). However, although it has been over twenty years since the first study on this line was published, to date we lack evidence of whether the performance of cognitive control tasks increases utilitarian decisions in moral dilemmas, which was the main goal of the current study.

Utilitarian decisions on moral dilemmas have been associated with more deliberative cognitive abilities (Baron et al., 2015; Moore et al., 2008). In addition, the induction of deliberative cognition has been shown to lead to an increase in utilitarian decisions (Li et al., 2019; Patil et al, 2020; Spears et al., 2018; Spears et al., 2021). For instance, traits like high working memory capacity (Moore et al., 2008), Actively Open-Minded Thinking, (AOT) and Cognitive Reflection (CRT; Baron et al., 2015) have all been associated with both higher cognition, as well as utilitarian decisions. Utilitarian decisions have also been induced through different tasks associated with

higher cognition. For instance, by reading dilemmas in difficult to read fonts compared to normal fonts as well (Spears et al., 2018); performing a cognitive reflection task before or after the dilemmas (Paxton et al., 2012); and giving feedback to incorrect answers after performing numerical task before responding dilemmas (Spears et al., 2021). In short, different manipulations and tasks performed before responding to dilemmas can in fact increase deliberative cognition, and thus increase utilitarian decisions, which suggests that using a cognitive control task before responding to dilemmas may also increase utilitarian decisions on moral dilemmas.

Though there have not been many, there are some studies showing that cognitive control tasks, such as the Stroop, can affect decision-making by decreasing intuitive responses and choices to stimuli. Kleiman et al., (2014) found that incongruent letter flanker trials (Eriksen & Eriksen, 1974) reduced both gender (Experiment 1) as well as racial stereotypes (Experiment 2), whereas two different Stroop tasks presented before a food stimulus (overt photo choice or implicit food Stroop) increased healthy choices for participants with higher body mass indices (Kleiman et al., 2016). While these experiments are not related to morality, they do show the possibility of using cognitive control tasks to induce deliberative cognition, and thus affect decisions. Timmons et al. (2018), on the other hand, found that when a task induces cognitive fatigue, such as having participants rewrite a text without using vowels compared just re-writing a text, utilitarian decisions are reduced. Trémolière et al. (2012) found a similar pattern where participants that had to memorize a complex dot matrix before responding to a moral dilemma made fewer utilitarian decisions. Therefore, there is a possibility that instead of inducing utilitarian decisions, high cognitive control could reduce them. In this study, Stroop and Flanker tasks were used with the aim of inducing higher cognition and thus increasing utilitarian decisions on moral dilemmas.

Participants in these experiments responded to moral dilemmas after being presented with a Cognitive Control task. One block of trials (Low Load) contained only congruent trials while the other (High Load) had 50% congruent and 50% incongruent trials. We hypothesized that participants would: H1) Make more utilitarian decisions after High Load blocks; H2) Take longer to make Utilitarian decisions after High Load blocks compared to Low Load blocks; and also, we hypothesized H3) that there would be an interaction between Load level (High/Low) and Dilemma type (Personal/Impersonal), where it is predicted that utilitarian decisions will increase more for personal dilemmas after High Load Blocks, while Impersonal Dilemmas will not be affected by Load Level. Mood was also measured to check for confounds of emotions on responses to dilemmas and/or performance on the Stroop (Timmons et al., 2018).

## 2. EXPERIMENT 1

### 2.1. Method

#### *2.1.1. Participants*

A total of 24 University of Granada students (16 male, 7 female, 1 unidentified; $M_{age}$ = 19.74) participated in exchange for course credit. All participants response times were within three standard deviations of the mean, though two participants were eliminated for having responses 3 standard deviations outside of the mean in errors in the Stroop task, leaving 22 participants in the analyses (14 male, 7 female, and 1 unidentified; $M_{age}$ = 19.64). All procedures were approved by the local Ethics Committee (#934/CEIH/2014).

#### *2.1.2 Materials*

In the Stroop task, the four words ‘green’, ‘red’, ‘blue’, and ‘yellow’, each colored in these four hues and displayed in uppercase at a 10 Arial font, served as stimuli for the task.

Dilemmas, on the other hand, included Personal and impersonal self-inevitable moral situations (Moore et al., 2008). *Personal* dilemmas are those in which the decision is to physically sacrifice a person with their own hands through pushing strangling, or firing a gun at close range, while *impersonal* dilemmas are usually caused by some instrumental act, such as pulling a lever or pushing a button to divert the danger towards the victim to be sacrificed. *Self* dilemmas are when the person who is deciding will be one of the people to die if they do not sacrifice one person. Finally, *inevitable* dilemmas are those where the person to be sacrificed will also be one of the five to die, even if the actor does not sacrifice them, therefore they will die regardless of whether the decision is utilitarian or deontological. The practice dilemmas in the task were neutral questions with no moral consequence such as choosing to add almonds to a brownie recipe instead of walnuts. The filler dilemmas are moral scenarios where participants had to decide whether they should lie on their taxes to buy a new car. The specific moral dilemmas were versions of *Burning Building*, *Rescue 911*; the practice dilemma was *Brownies*; while the fillers were *Taxes*, and *Stock Tip*[14].

[14] Before reporting any results, it would be prudent to indicate that there was an unforeseen problem in the programming of the experiment. While it was our intention to have dilemmas completely randomized, it appears that they followed a sequential pattern. That is, if a participant first saw the personal version of Rescue 911 after the first block of High Load, the very next dilemma that s/he saw was the Impersonal version of Rescue 911 from Block B after the first block of Low Load. While this is not the best scenario, the data is still useful because the dilemmas and blocks were counterbalanced, and the very first dilemma was random.

In order to be certain that participants had carefully read each dilemma before responding, they had to answer one content question for each dilemma. For example, the question for the dilemma *rescue 911* asked: *What part of the helicopter was the equipment bolted onto?*, with a correct answer of 'floor'. All the study materials can be found in Appendix A.3.1.

For the mood ratings, participants indicated their mood on a 7-point scale (from 1 = very sad, to 4 = neither sad nor happy, and 7 = happy), which was previously used in Spears et al. (2018) and was based on the scale from (Alter et al., 2007).

*2.1.3 Procedure*

Participants read and signed an informed consent before reading detailed instructions to the experiment. Then they responded to the first mood scale, then had an opportunity to practice both the Stroop task (12 trials, 50% incongruent) and a neutral practice dilemma. Then they first responded to a Stroop block, either starting with the High or Low Load depending on the counterbalance, followed by a single dilemma. Afterwards, they responded to another mood scale, then repeated the cycle, alternating between Low and High Load until they responded to all six dilemmas.

In all the Stroop blocks, the sequence of events was the following (Figure 3.1). In each trial, a fixation cross (+; 1.15º gav) was presented in the middle of the screen for 1000 ms, and then the colored word (1.15º gav) appeared at the same location for up to 2000 ms. Participants had 2000 ms to respond to the hue color using the *d, f ,j,* and *k* keys on the keyboard. Once a participant responded to the word, the trial immediately ended, and a new fixation cross was presented. In total, there were 144 Stroop trials in the whole experiment, 72 trials (three blocks of 24) in the High Load blocks (36 congruent, 36 incongruent) and 72 in the Low Load blocks.

The dilemmas appeared at the center of the screen and lasted until the participant responded. Due to limits in the design capabilities of the PsychoPy software to randomize and counterbalance the order of dilemmas, these were split into two sets: A) Personal *Rescue* 911, Impersonal *Burning Building*, and filler *Taxes*; and B) Impersonal *Rescue 911*, Personal *Burning Building*, and the filler *Stock Tip*. Half of participants received the A order first before completing the B set, whereas the rest completed first set B and then set A. The session took between thirty and forty-five minutes for participants to complete.

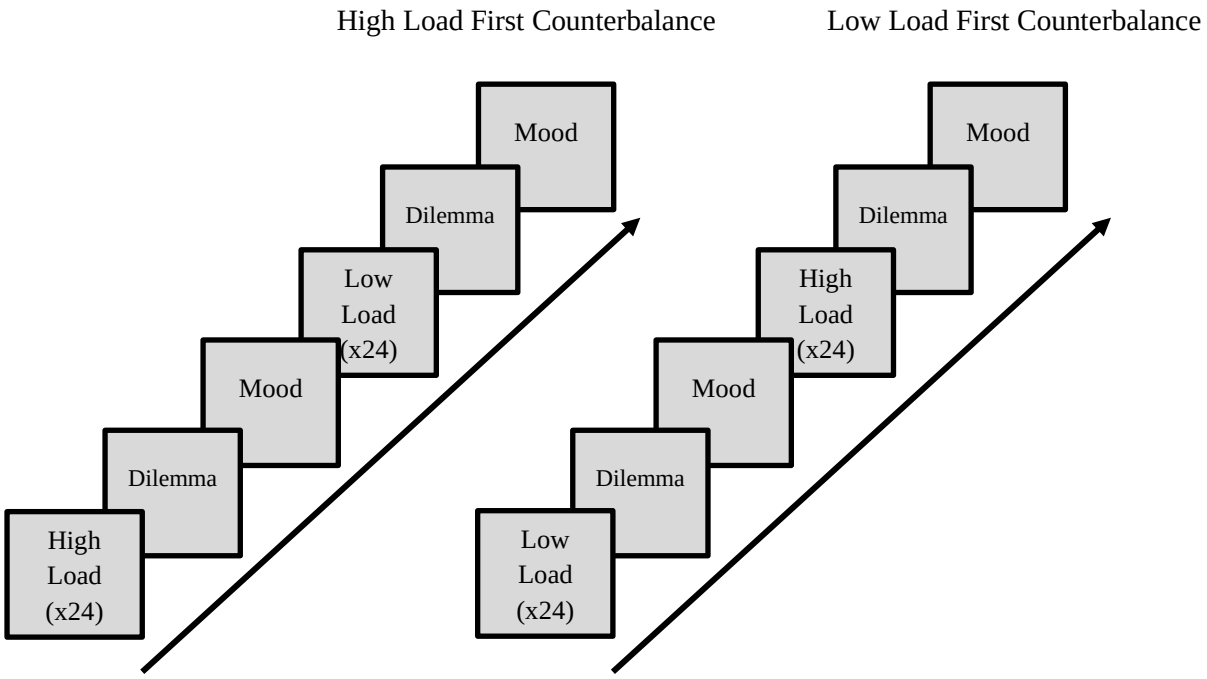


Figure 3.1. Flow chart of the experimental procedure. Half of participants saw High Load blocks first, while the other half responded to Low Load cognitive control blocks first. After the first block of cognitive control, participants responded to a dilemma, followed by the self-reported mood scale. Afterwards, they responded to the opposite block of cognitive control. Therefore, if a participant saw a block of High Load first, the second block was Low load. Blocks repeated and alternated in load level 3 times for each High and Low Load.

## 2.5. Results

### *2.5.1. Manipulation Check*

First, we analyzed response time and percentage of errors across Stroop trials. A one-way repeated measures ANOVA was performed comparing the errors of High and Low load blocks, which revealed that there was no significant difference in errors between them, $F$ (1, 21) = 3.34, $p$ = .079, $\eta_p^2$ = .14. However, a similar analysis did find a significant difference in RTs (response times) between High and Low load blocks, $F$ (1, 21) = 38.05, $p < .001$, $\eta_p^2$ = .64, where participants took longer to respond to High Load Stroop blocks compared to Low Load blocks (RTs: High Load $M$ = 0.75 ms, SE = 0.02; Low Load $M$ = 0.65 ms, SE = 0.02).

A further analysis of Stroop trials collapsed across all High Load blocks (blocks with 50% Congruent trials). This analysis revealed a significant effect of Congruency on errors, $F$ (1, 21) = 7.74, $p$ = .011, $\eta_p^2$ = .27, where Incongruent trials had more errors compared to Congruent ones (Number of errors: Incongruent $M$ = 1.33 errors, SEM = 0.22; Congruent $M$ = 0.58 errors, SEM = .13). Further, a repeated measures ANOVA of Response Times (RTs) comparing congruent and incongruent trials in High Load blocks revealed a significant effect of Congruency on RTs, $F$ (1, 21) = 47.89, $p < .001$, $\eta_p^2$ = .70, where participants were faster to respond to congruent trials opposed to incongruent trials (Congruent $M$ = 679.59 ms, SEM = 23.73; Incongruent $m$ = 824.35 ms, SEM = 35.08). To ensure that the order of the counterbalance did not affect errors, a mixed measures ANOVA was done with Load Level (High x Low), as the within-subject factor, and Order as between-subject factor (High Load First x Low Load First), which found no interaction between both, $F$ (1, 20) = 0.083, $p$ = .78, $\eta_p^2$ = .004. This was similarly found when analyzing Trial Congruency (Congruent x Incongruent) as within subject factors and Order (High Load First x

Low Load First), $F$ (1, 20) < 0.001, $p$ = .98, $\eta_p^2$ < .001. We obtained similar results with RTs on Load Level (High x Low) as within subject factors and Order (High Load First x Low Load First) as the between subject factor, $F$ (1, 20) = 2.66, $p$ = .12, $\eta_p^2$ = .008, and RTs on Trials (Incongruent x Congruent) as the within subject factor and Order (High Load First x Low Load First), $F$ (1, 20) = 0.089, $p$ = .77, $\eta_p^2$ = .004. Therefore, order had neither an effect on errors nor response times in the Stroop task.

*2.5.2. Mood*

To test Mood, we ran a mixed ANOVA on Timing of assessment (7 Levels: Before, second, third, fourth, fifth, sixth, and end) as within subjects factor, and Order (High Load First versus Low Load First) as the between subjects factor. There was a main effect of Timing, $F$ (6, 120) = 6.97, $p$ < .001, $\eta_p^2$ = .26, where the Mood score decreased as the experiment progressed (A Bonferroni post hoc test revealed that there were significant differences between the first Mood rating and the fourth ($p$ < .001,), fifth ($p$ < .001), sixth ($p$ < .001), and final mood rating ($p$ < .001). There were no other significant differences (smallest $p$ = .44), indicating that after the first mood rating, participants mood scores dropped and remained low. There were no interactions between Mood and Order, $F$ (6, 120) = .61, $p$ < .72, $\eta_p^2$ = .03. Therefore, while Mood decreased throughout the experiment, this was found regardless of whether participants responded to High Load or Low Load blocks first.

*2.5.3. Cognitive control and moral dilemmas*

First, to test H1 (utilitarian responses to dilemmas will increase after High Load compared to Low Load blocks) we ran a 2 (Load: Low vs High) x 2 (Dilemma Type: Personal dilemmas vs. Impersonal dilemmas) on responses to dilemmas. These results revealed a main effect of Dilemma

type, $F$ (1, 21) = 16.82, $p < .001$, $\eta_p^2 = .445$, where participants rated acceptability of impersonal dilemmas ($M$ = 4.27, $SEM$ = .22) higher than that of personal dilemmas ($M$ = 3.30, $SEM$ = .24). However, there were no further main effects or interactions, largest $p$ = .125.

The next analysis was another 2 Load (High/Low) x 2 Dilemma Type (Personal/Impersonal) repeated measures ANOVA on response times to dilemmas (RTs). This analysis revealed neither any main effects nor interactions, largest $F$ (1, 21) = .67 $p$ = .417. Finally, a mixed measures ANOVA was done to determine whether the order of cognitive control blocks had an effect on responses to dilemmas, with Load Level (High vs. Low) x Dilemma Type (Personal vs. Impersonal) as within subject factors, and Order (High Low First vs. Low Load First) as between subject factors. This analysis revealed no interactions with Order (largest $F$ (1, 20) = 2.20).

**2.6. Discussion**

The aim of this experiment was to induce utilitarian decisions on moral dilemmas after a cognitive control task. Participants responded to several blocks of Stroop trials where some blocks were 100% congruent alternating with blocks containing 50% of incongruent trials. In between each block, participants responded to a single moral dilemma.

Firstly, analysis of errors and response times between congruent and incongruent Stroop trials, and blocks showed that the Stroop manipulation was effective, with participants performing worse and slower on incongruent than on congruent trails in the High Load Blocks. However, regarding the relation between cognitive control and utilitarianism, utilitarian decisions did not increase after High Load blocks of the Stroop task. Thus, we cannot reject the null hypothesis, H1, where utilitarian decisions would not increase after responding to High Load blocks compared to

responses on dilemmas on Low Load blocks. One explanation for this might be that the Stroop task, regardless of load level, can cause cognitive fatigue (Timmons et al., 2018), which may hinder utilitarian deliberative reasoning. Therefore, the next experiment used a Flanker task, with binary responses, which may cause cognitive fatigue to a lower extent (Timmons et al., 2018).

## 3. EXPERIMENT 2

The Flanker task has been shown to lower participants' gender and race stereotypes after completing incongruent trials (Kleiman et al., 2014). Further, its performance has been shown to activate similar areas of the brain as utilitarian decisions do (Greene, et al., 2001; Mansouri et al., 2009). The main objective of Experiment 2 was to examine whether participant's utilitarian decisions increased after completing a Flanker task. As we were not interested in comparing the effect on different types of dilemmas, we decided to simplify the design by removing dilemma type and Response Time analyses on dilemmas, as we found not significant effects or interaction in the previous experiment. We further reduced the number of mood measurements and only measured mood before and after all the blocks.

Similar to the previous experiment, utilitarian decisions were predicted to increase after High Load Flanker blocks compared to Low Load flanker blocks.

### 3.1. Methods

#### *3.1.1. Participants*

Participants were Psychology students from the University of Granada (N = 92)[15] who completed the experiment in a laboratory for course credit. From the initial sample, there were 17 participants eliminated (13 for answering at least three content questions incorrectly, 3 participants for having errors 3 standard deviations outside of the mean on the flanker task, and 1 participant for having flanker response times 3 standard deviations outside the mean). This left an N = 75 ($M_{age}$ = 19.84, 65% female).

*3.1.2 Materials*

The flanker task used the letters 'S' and 'H'. Each trial had five letters in a row, with congruent trials being 'SSSSS' or 'HHHHH'. and incongruent trials being 'SSHSS' or 'HHSHH' (see Kleiman et al., 2013).

All dilemmas were personal. The specific dilemmas used were *Bus Crash, Space Station, Nuclear reactor, Sky Scrapper, Burning Building,* and *Rescue 911*. As in Experiment 1 (the Stroop experiment) Dilemmas were separated into two blocks: A, which contained *Rescue 911, Bus Crash,* and *Space Station*; and B, which contained *Nuclear reactor, Sky Scrapper, and Burning Building*. All dilemmas were translations taken from Carmona-Perera et al. (2013) except *Burning Building* and *Rescue 911*, which were taken from Spears et al. (2018). Participants responded to the question: *Do you think that you should sacrifice one person in order to save yourself and the*

[15] The differences between sample sizes in the first and second experiment is due to the preliminary results in the first experiment showing that results were far from significant, while preliminary analyses on the second experiment were significant at 24 participants, so more participants were included to insure this pattern.

*rest of the passengers'*, using a scale from 1 – 6, where 1 = "definitely No", and 6 = "definitely Yes". All dilemmas can be found in Appendix A.3.1.

The Mood scale is the same as Experiment 1.

## 3.2. Procedure

The pre-experimental and practice session stages for this experiment were the same as in Experiment 1. In each trial of the Flanker task, there was a fixation cross (+, 1.15 GAV) that was presented for 1000 ms, and next the flanker stimuli, which were presented for 2000 ms, during which participants could respond by pressing the H and S keys. The timings and display of the dilemmas were the same as in the previous experiment. The experimental phase of the session also followed the same procedure, though there were 36[16] trials in each Cognitive control block. In total, there were 216 flanker trials. The whole experiment took between 15 and 20 minutes to complete.

## 3.3. Results

### *3.3.1. Flanker Analysis*

Repeated Measure ANOVAs for errors and response times for Low Load blocks (100% Congruent trials) and High Load Blocks (50% Congruent) collapsed across blocks were also done. Participants made more errors in the High Load blocks compared to the Low Load blocks, $F$ (1, 74) = 19.10, $p < .001$, $\eta_p^2 = .21$ (Number of errors: $M_{HighLoad}$ = 4.12 errors, SEM = .28; $M_{LowLoad}$ = 2.89 errors, SEM = .25), and took longer to respond in High Load blocks compared to Low Load

[16] The number of trials increased on the flanker task to try to equalize the amount of time participants spent completing the Stroop blocks as well as flanker.

blocks, $F$ (1, 74) = 220.66, $p < .001$, $\eta_p^2 = .75$ ($M_{HighLoad}$ = 463.39 ms, SEM = 4.71; $M_{LowLoad}$ = 425.66 ms, SEM = 4.90).

Flanker trials were analyzed using a one-way repeated measures analysis of variance (ANOVA). First, High Load block trials were collapsed across blocks, and then mean errors and RT were compared. As expected, participants made significantly more errors in incongruent trials, $F$ (1, 74) = 89.75, $p < .001$, $\eta_p^2 = .55$ (Number of errors: $M_{Incongruent}$ = 3.23 errors, SEM = .24; $M_{Congruent}$ = .89 errors, SEM = .11) and took significantly longer to respond, $F$ (1, 74) = 160.78, $p < .001$, $\eta_p^2 = .69$ ($M_{Incongruent}$ = 482.82 ms, SEM = 4.94; $M_{Congruent}$ = 444.91 ms, SEM = 4.91). This indicates that, in the High Load block, a flanker congruency effect was induced.

To ensure that there were not any order effects on Errors in the flanker task, a Mixed design ANOVA was done with Load Level (High x Low) as the within subject factor and Order (High Load First x Low Load First) as the between subject factor, which revealed no interaction, $F$ (1, 73) = 0.06, $p = .80$, $\eta_p^2 < .001$. This same analysis, done with Trial Congruency (Incongruent x Congruent) as the within subject factor and Order (High Load First x Low Load First) as the between subject factor, $F$ (1, 73) = 1.22, $p = .27$, $\eta_p^2 = .001$, also yielded no interaction between Trial Congruency and Order. These analyses were done on RTs, with Load Level (High x Low) as the within subjects factor and Order (High Load First x Low Load First) as the between subjects factor, $F$ (1, 73) = 0.34, $p = .56$, $\eta_p^2 = .006$, where there was no interaction between errors and order. Finally, there was also no significant interactions between Congruency (Incongruent x Congruent) as the within subjects factor and Order (High Load First x Low Load First) as the between subjects factor, $F$ (1, 73) = 0.45, $p = .51$, $\eta_p^2 = .005$. Therefore, order did not have an effect on errors or responses times.

*3.3.3. Mood*

A mixed design ANOVA, with Mood Measurement (Before versus After) as the within-subject factor, and Order (High Low First versus Low Load First) as between subject factors was done to analyze mood scores. This analysis revealed a main effect of Mood Measurement, $F$ (1, 60) = 55.43, $p < .001$, $\eta_p^2 = .48$, where the Mood before the experiment began was higher than the mood at the end. However, there was neither a main effect of Order, nor an interaction.

*3.3.4. Dilemma Analysis*

Responses to dilemmas were averaged across blocks before conducting a repeated measures ANOVA, with Load Level as the factor (High Load x Low Load) on dilemmas responses. This analysis revealed that participants made significantly more utilitarian decisions after the Low Load blocks compared to responses after High Load, $F$ (1, 74) = 5.16, $p = .026$, $\eta_p^2 = .07$ ($M_{High\ Load} = 3.01$, SEM = 1.45; $M_{Low\ Load} = 3.18$, SEM = .15). Further, a mixed design ANOVA, with Load Level (High Load x Low Load) as the within subjects factor, and Order (High Load First x Low Load First) as the between subjects factor, did not reveal any interactions between Load Level and Order, $F$ (1, 73) = 3.55, $p = .064$.

**3.4. Discussion**

In this second experiment, we attempted to induce more deliberate cognition through the Flanker task, which aimed to prompt cognitive control and thus to increase utilitarian decisions on moral dilemmas. The analysis of errors and response times on the Flanker tasks suggested that more cognitive control was necessary to respond correctly in the High Load blocks. However, a corresponding increase in utilitarian decisions after High Load blocks of Flankers was not found, giving no support to H1. Instead, increased load significantly reduced utilitarian decisions on moral

dilemmas, which suggests that High Load may have induced cognitive fatigue in participants and thus decrease utilitarian decisions.

## 4. GENERAL DISCUSSION

In this study, we set out to induce deliberative processing using two cognitive control tasks, Stroop (Stroop, 1935) and Flanker (Eriksen & Eriksen, 1974), under the expectation that increased cognitive control would increase utilitarian decisions on moral dilemmas.

Manipulation checks showed both Stroop and Flanker effects, where participants committed more errors and had slower response times in High Load blocks compared to low. The order of Load Levels of blocks and dilemmas were counterbalanced across tasks and dilemmas blocks, and there were no order effects on errors or responses times in the tasks nor on responses to the dilemmas. In the first experiment, using the Stroop task, while the incongruent trials in the Stroop task did increase cognitive control, this did not have an effect on responses to moral dilemmas. However, after participants responded to a Flanker Task with a high proportion of Incongruent trials, their responses to dilemmas seemed to be affected. Contrary to our initial hypotheses, participants made fewer utilitarian decisions after the High Load cognitive control blocks.

The reason why utilitarian decisions decreased after High Load compared to Low could be due to cognitive fatigue. Bialek and De Neys (2017) found that participants made fewer utilitarian decisions when responding to moral dilemmas after memorizing and before recalling high load dot matrix compared to low load. That is, when participants needed to remember more complicated patterns which they needed to recall later, they made fewer utilitarian decisions. However, Bialek

and De Neys (2017) had participants read the first part of the dilemma, then they had to memorize a dot matrix, before reading the second part and responding to the dilemma, and finally selecting the correct answer. The authors' intention was to overload working memory, which is positively associated with utilitarian decisions (Moore et al., 2008), in order to reduce the utilitarian decisions. Further, by splitting the dilemma and the choice with a difficult memory task, Bialek and De Neys (2017) disrupted the decision-making process. Using a sequential task, Timmons et al. (2018) showed that, after being made to rewrite a paragraph without using the letter 'e' or use the 'spacebar', participants made fewer utilitarian responses when compared to those participants who participants who had no restraints when rewriting the paragraph. However, the tasks used by Timmons et al. (2018) were intentionally taxing cognitive resources, with the intention to cognitively exhaust participants. The intention of the experiments reported in this study was not to overload or disrupt cognition in the midst of deciding such as the dot matrix memorization task (Bialek, & De Neys, 2017), but to increase cognitive control. However, we cannot discard an effect of fatigue induction in our Experiment 2.

Lavie's Load Theory (Lavie & Dalton, 2014) argues that high load cognitive control tasks would focus attention onto the difficult task and depletes resources, which cannot be applied to other tasks. That is, a person would be so concentrated on a difficult task, that any subsequent task would get less attentional resources. In this interpretation, it could be that participants were more concentrated and focusing their cognitive resources on doing the Stroop or Flanker task well and then relaxed mentally, thus they did not deliberate on their decisions on the moral dilemmas which came after each Stroop or Flanker block. While this is mostly speculative, the null effect in the Stroop experiment is congruent with this interpretation considering how much more demanding a four-color Stroop task is compared to the binary Flanker. In fact, Rauch and Schmitt (2009) found

that cognitive fatigue started affecting performance on Stroop tasks within ~15 minutes of doing all congruent trials, which the authors explained was due to the complexity of the task and the need to maintain the colors in working memory. While some tasks can induce a more deliberative mindset, such as the CRT, or BNT (Spears et al., 2020), others may tire participants.

### *4.1 Limitations*

One limitation in this study is the use of traditional moral dilemmas. Recent studies are using congruent and incongruent dilemmas (Conway et al., 2013), which are used to analyze whether decisions are increased deontological, increased utilitarian, or increased harm, which could help determine if increased cognitive load decreases endorsement of utilitarian decisions or increases endorsements of deontological decisions. Further, future studies using the Stroop task should compare responses to dilemmas after a two-color Stroop with a four-color Stroop, which will help the understanding of the complexity of a cognitive control tasks on moral dilemmas.

In conclusion, across two experiments, High Load cognitive control tasks did not increase utilitarian decisions on moral dilemmas. In fact, in at least one case, the effect may be the opposite, where more incongruent cognitive control tasks might reduce utilitarian decisions. This data leans towards evidence for the moral fatigue effect (Timmons et al., 2018).

# Chapter 4

# Psychopathy, Reflection, and Moral Dilemmas

# Study 4:

# Cognitive reflection and Callous affect: two independent factors related to utilitarian decisions on moral dilemmas

Based on the publication:

Spears, D., Okan, Y., Cándido, A., & González, F. (2014). Cognitive reflection and callous affect: Two independent factors related to utilitarian decisions on moral dilemmas. In A. Acosta, J. L. Megías, & J. Lupiáñez (Eds.), *Avances en el estudio de la motivación y la emoción* (pp. 128–134). AME.

# FOREWARD

There is a question on whether utilitarian decisions on moral dilemmas come from a deliberative thinking (Greene et al., 2001; Li et al, 2018; Paxton et al., 2012; Spears et al., 2018; Spears et al., 2021) or from emotional blunting – the weakening of emotional responses to a stimulus (Bartels et al., 2009; Koenigs et al., 2008).

Recent evidence (Conway et al., 2018; Patil et al., 2014) has found that emotional blunting, as measured by psychopathy or antisocial traits, might not be a straightforward explanation. Conway et al. (2018) found that participants with antisocial traits were just as likely to sacrifice one person's life to stop five other people from being hurt (which is both against utilitarianism as well as deontology), as they were to make utilitarian decisions. This suggests the possibility that participants with psychopathic or antisocial traits do not make utilitarian decisions for the greater good, but instead make harmful decisions.

The study reported below attempts to broaden this understanding on the differences of utilitarian decisions prompted by deliberative thinking or by blunted social emotions, and to analyze the relationship (or lack of it) between them.

## 1. INTRODUCTION

When a person decides to take another person's life, are they being rational, cold, or both? Utilitarian responses in sacrificial moral dilemmas (i.e., agreeing to the death of one person in order to save several lives) are more frequent when prepotent emotional responses and intuitions are inhibited (Greene et al., 2001; Paxton et al., 2012). Greene et al. (2001) showed that high conflict personal dilemmas (Koenigs et al., 2007) induce an emotional response in the person making the decision. Impersonal dilemmas (Standard Trolley, and Standard Fumes) are those where the decider does not have to do direct harm in order to save others' lives; as a consequence, the action to save five other people is less emotionally salient than in personal dilemmas (Greene et al., 2001; Greene et al., 2004). For instance, a person only has to pull a lever that changes the track of a train out of control that will hit and kill another person saving five people on the train from death. In contrast, personal dilemmas are those where the decider must kill another person directly in order to save more lives – that is, the agent must physically kill one person to save several others, which also "engages" more intense emotional responses (Greene et al., 2001; see example below). These personal dilemmas are separated into two sub-categories: High and Low conflict (Koenigs et al., 2007). Personal low-conflict dilemmas (e.g., Plan Crash, Transplant) have shown to have nearly 100% deontological decisions (i.e., disagreeing with the proposed action), while high-conflict dilemmas (e.g., Footbridge, Crying Baby) have more variations in their decisions, as they are thought to produce a conflict or competition between aggregate welfare and harm to others (Koenigs et al., 2007). An example of a dilemma, taken from the personal high-conflict category, the '*'Footbridge'* dilemma goes:

> A runaway trolley is heading down the tracks towards= five workmen who will be killed if the trolley proceeds on its present course. You are on a footbridge over the tracks, in between the approaching trolley and the five workmen. Next to you on this footbridge is a stranger who happens to be very large.
>
> The only way to save the lives of the five workmen is to push this stranger off the bridge and onto the tracks below where his large body will stop the trolley. The stranger will die if you do this, but the five workmen will be saved.
>
> Would you push the stranger on to the tracks in order to save the five workmen?

In support for this distinction, in Koenigs et al. (2007) study it was shown that the areas of the brain activated when participants read the dilemmas had been previously associated with emotional processing. Greene et al. (2004) showed that, when making utilitarian decisions in difficult personal dilemmas (not necessarily high conflict), the anterior cingulate cortex (ACC) and dorsolateral prefrontal cortex (dlPFC), areas of the brain previously associated with cognitive control, were activated.

Regarding the effects of cognitive control and deliberative thinking on moral decision-making, Paxton et al. (2012) gave participants the Cognitive Reflection Test (CRT; Frederick, 2005), a numeric task where the intuitive answer, prompted by the scenario, to questions are wrong, being necessary to override them by reflection in order to reach the right response. The most famous question from this task is the 'bat and ball' problem: '*A bat and a ball cost $1.10 together. The bat costs $1.00 more than the ball. How much does the ball cost?*'. The intuitively answer to this question is $.10, which is incorrect since the total would then be $1.20 (ball = $.10, bat = $1.00 + ball = 1.10; so, $.10 + $1.10 = $1.20). Participants who completed the CRT before evaluating the high conflict personal dilemmas and answered questions correctly gave more

utilitarian responses. This test was used to assess participants' ability to inhibit intuitive responses and to examine whether any relation existed between this ability and their tendency to make utilitarian decisions in moral dilemmas.

However, as previously commented, cognitive control is not the only factor that leads to utilitarian responses. Bartels et al. (2011) demonstrated that people who have antisocial personality traits (the authors referred to as psychopathic traits) measured by the Self-Reported Psychopathy questionnaire (SRP-III; Paulhus & Kelly, 2002; being cold or calloused), tend to make more utilitarian choices. Further, Kahane et al. (2015) found that participants with immoral outlooks, willingness to break business ethics rules, as well as psychopathic traits also support utilitarian decisions on moral dilemmas. In both studies, the authors interpreted this data as evidence that utilitarian decisions do not necessarily stem from deliberative cognition. Other studies have shown that utilitarian decisions can come from a lack of empathetic concern in general, including patients with lesions in the ventromedial prefrontal cortex (Ciaramelli et al., 2007; Koenigs et al., 2007; Moretto et al., 2010); reduced inhibitions in social cognition through consuming alcohol (Duke et al., 2014); or alexithymia (Patil et al., 2014). Wiech et al. (2013) found that high scores on the Psychoticism sub-scale of the Eysenck Personality Questionnaire (Eysenck & Eysenck, 1991)—which measures diminished empathetic concern and aggression-- and Need for Cognition (Cacioppo et al., 1996), a measure of a person enjoyment and pursue effortful cognitive activity—both positively correlated with utilitarian decisions, but not each other.

Similarly, the study presented here considers psychopathy as a collection of anti-social traits of (Paulhus, & Kelly, 2002): calloused affect (CA), interpersonal manipulation (IPM), erratic lifestyle (ELS), and Anti-Social Behavior (ASB). CA tests how cold and calculating a person is; a participant that scores high on CA would strongly agree with the statement: 'people sometimes

say I'm cold hearted'. IPM tests the willingness of a person to use another and not care about their reactions and would strongly agree to the statement: 'it's fun to see how far you can push people before they get upset'. ELS tests how willing a person is to participate in risky behaviors and would strongly agree with the statement: 'I rarely follow the rules'. Finally, ASB measures the willingness a person has in seeing or causing harm to others and would say that they would 'not be tormented seeing an animal injured'. In this study, apart from total psychopathy traits, the subscale CA was measured since this subscale focuses on the emotionally blunted aspect of psychopathy, which this study is interested in exploring.

Is it the case that people who are cold and calloused in their moral decisions are also the ones that are able to inhibit intuition? In our study, we measured both participants' CRT performance and psychopathic personality traits to relate these two different factors to utilitarian responses. We included an analysis of the subscale CA associated with the anti-social personality traits, which was lacking in previous studies. We expected to find significant positive correlations between both the CRT, the SRP-III score, CA subscale score, and utilitarian responses, but not between the correct answers on the CRT and the SRP-III or CA subscale. In this study, we predicted that participants who score higher on psychopath traits (SRP-III) and CA subscale will make utilitarian decisions; that participants who score higher on the CRT will make more utilitarian decisions on moral dilemmas; though, participants who answer more questions correctly on moral dilemmas will not score higher in psychopathy or the CA subscale.

## 2. Method

### *2.1.Participants*

Ninety-five undergraduate students (64 female) from the University of Granada participated in the experiment in exchange for class credits. Participants completed the experiment in individual computers using the Web survey Unipark (www.unipark.de). All dilemmas and questionnaires were translated into Spanish.

*2.2. Dilemmas*

The sacrificial dilemmas were selected for being either non-moral, impersonal, (Greene et al., 2001), personal high-conflict, and personal low-conflict (Greene et al., 2004; Koenigs et al., 2007). Non moral dilemmas were used as a control. The two impersonal dilemmas were *Trolley*, and *Fumes*; the personal low-conflict dilemmas were *Transplant*, and *Footbridge*; and the two personal high-conflict dilemmas were *Plane Crash*; and *Crying baby*. After reading the dilemmas, participants were asked if they should harm one person to save five by choosing a response option ranging from 1 (absolutely no) to 6 (absolutely yes) to each dilemma. Higher scores indicate greater preference for the utilitarian solution. All dilemmas can be found in Appendix D.1.

*2.3.Questionnaires*

After responding to the dilemmas, participants were presented with the CRT (Frederick, 2005) and the SRP-III (Paulhus, & Kelly, 2002). The exact questions can be found in Appendix D.2.

The CRT is a series of three questions where, in order to answer correctly, one must inhibit an intuitive response. Correct answers for each question were scored as 1 and incorrect responses were scored as 0, with a maximum total score of 3. The three questions used can be found in Appendix D.3.

The Psychopathy scale (SRP-III) consisted of answers to a series of questions (Bartels et al., 2011; a shortened and revised version of the scale by Paulhus, & Kelly, 2002). It is also made up of several subscales (commented above: CA, IPM, ELS and ABS) to test the level of antisocial personality traits a person may have.

### 3. Procedure

Participants completed the moral dilemmas task first, after been randomly allocated to one of two orders of presentation: 1) neutral, impersonal, low-conflict, and high-conflict; or, 2) high-conflict, low-conflict, impersonal, and neutral, following the ordered design studied and described by Wiegmann et al. (2012). Studying the ordered effects of dilemmas, they found that more emotionally salient dilemmas reduce utilitarian responses on subsequent dilemmas that were less emotionally salient, but not vice versa. That is, if one dilemma leads to a stronger emotional response it will affect the responses to the next dilemmas by reducing utilitarian responses. They were free to read the text of each dilemma at their own pace and after clicking on 'continue' they viewed the response scale (1 - 3 = "No"; 4 - 6 = "Yes") for each dilemma on a separate screen.

Finally, participants were given the CRT and the Psychopathy scale, in that order.

### 4. Results

The scores of the three types of dilemmas were averaged across dilemma type (Impersonal, Personal Low Conflict, and Personal High Conflict) and were submitted to repeated measures ANOVA. This analysis revealed a significant main effect of type of dilemmas, $F$ (2, 188) = 179.24, $p < .001$, $\eta_p^2 = .66$, where participants made more utilitarian decisions on Impersonal dilemmas ($M$ = 4.41, S.E. = .11), compared to both Personal Low Conflict ($M$ = 2.13, S.E. = .10) and Personal High Conflict ($M$ = 2.79, S.E. = .13).

Table 4.1: Correlations between dilemmas, CRT, anti-social personality total score and CA subscale.

Moral dilemmas: I = impersonal, LC-P: low-conflict personal; HC-P: high-conflict personal

| Correlations | | | | |
|---|---|---|---|---|
| | | CRT | Psychopathy Total | CA |
| **Trolley (I)** | | | | |
| | Pearson | -.095 | .123 | .173 |
| | Sig | .361 | .234 | .093 |
| **Fumes (I)** | | | | |
| | Pearson | -.014 | .187 | **.212*** |
| | Sig | .893 | .070 | .039 |
| **Transplant (LC-P)** | Pearson | -.017 | -.017 | -.004 |
| | Sig | .871 | .870 | .971 |
| **Footbridge (LC-P)** | Pearson | .150 | .198 | **.242*** |
| | Sig | .147 | .054 | .018 |
| **Plane Crash (HC-P)** | Pearson | -.100 | .129 | .088 |
| | Sig | .337 | .213 | .397 |
| **Crying Baby (HC-P)** | Pearson | **.210*** | **.466**** | **.471**** |
| | Sig | .04 | < .001 | < .001 |

Note: * = Correlation is significant at the .05 level (two tailed)

* * = Correlation is significant at the .01 level (two tailed)

Average utilitarian responses to Personal High Conflict dilemmas were also higher compared to Personal Low Conflict. Pairwise comparisons (Bonferroni's correction) confirmed significant

differences between Impersonal and Personal Low Conflict ($p < .001$), and Personal High Conflict ($p < .001$), as well as between Personal Low Conflict and Personal High Conflict ( $p < .001$).

In terms of correlations between responses to dilemmas and CRT and Psychopathy scores (see Table 4.1), the response to one dilemma in particular ('crying baby') correlated with the CRT score, the overall psychopathy score, as well as with the (CA) subscale. Indeed, CA score correlated with utilitarian responses in one of each type of dilemma. Interestingly (see Table 4.2), the correlation between the CRT and psychopathy scale, or between the CRT and the CA subscale were not significant.

Table 4.2: Correlations between CRT anti-social personality (total) and CA subscale

| **Correlations** | | |
|---|---|---|
| | | CRT |
| **Psychopathy Total** | Pearson | .139 |
| | Sig | .181 |
| **CA** | Pearson | .094 |
| | Sig | .366 |

Note: * = Correlation is significant at the .05 level (two tailed)

* * = Correlation is significant at the .01 level (two tailed)

## 5. Discussion

The differences in utilitarian responses between impersonal, personal high-conflict, and personal low-conflict dilemmas followed the same pattern as past studies (Greene et al., 2001; Greene et al., 2004). Responses to the low-conflict personal dilemmas had a high proportion of deontological choices not to sacrifice ("No" to the proposed action), while the high-conflict personal had more utilitarian responses to sacrifice ("Yes"), and Impersonal dilemmas had the highest responses of utilitarian decisions.

Regarding cognitive reflection (CRT), the more they were capable of inhibiting intuitive responses the more a participant agreed with the utilitarian action in the crying baby dilemma. This result also agrees with previous research. Paxton et al. (2012) found that individuals who reflected more made more utilitarian judgments. In addition, they reported that CRT was associated with increased utilitarian judgment in a follow-up study.

Moreover, participants who also scored higher on both the psychopathy scale and CA, also tended to choose more utilitarian responses as well. While Bartels and Pizarro (2011) only analyzed the total score in the anti-social personality scale, we further examined the relationship between utilitarian responses and CA score, which correlated with utilitarian responses in one of each type of dilemma. Though, there were not any significant correlations between higher scores on the CRT with higher scores on both the psychopathy scale in total or the CA subscale. While these correlations do not show a causal influence, the fact that there was no correlation between the CRT and Psychopathy (or the CA subscale) scores suggests that there may be two different factors that lead to utilitarian decisions. A person that has psychopathic traits, and/or is colder and calculating, is not necessarily better at inhibiting intuition. On the other hand, cognitive reflection

does not seem to be related to impairments in feeling emotions towards another person when they make drastic decisions.

This data, instead, adds credence to the Dual-Process model of morality (Greene et al., 2004). The Dual-Process model of morality argues that utilitarian responses on moral dilemmas will either be deontological-intuitive/emotional or utilitarian-rational/deliberative. However, another interpretation is that utilitarian decisions are made, not by rational consideration of the costs and benefits of the sacrificial choice, but rather by reduced emotions (Bartels et al., 2011; Kahane et al., 2015). The data presented here furthers the interpretation that utilitarian decisions are, at least sometimes, rational.

Weich et al. (2013) found similar results where higher scores on the Psychoticism scale (Eysenck, & Eysenck, 1991) and NFC (Cacioppo & Petty, 1982) both positively correlated with utilitarian decisions, but not with each other. The authors suggest that this is a weakness or fault in the dual-process model of morality. Since the dual-process of morality argues that utilitarian decisions are the rational (Greene et al., 2004), and reflective decisions (Paxton et al., 2012), then any utilitarian decision that is not made under rational consideration weakens the dual-process model. However, Conway et al. (2018) found that people with anti-social traits are not necessarily making utilitarian decisions, but rather choosing to do harm directly by supporting such acts as choosing to test chemicals on animals that can lead to discomfort or pain to create beauty products that make people feel better.

As a conclusion, the present study add evidence suggesting that, while people with anti-social traits may choose utilitarian decisions, that does not mean that they are supporting the utilitarian outcomes. Instead, they may be more interested in the deontological violation.

# Chapter 5

# Morality, Politics, and Deliberation

# Study 5:

# Political and Economic beliefs, deliberative thinking, and moral decisions

## 1. INTRODUCTION

Deciding what is ‘good’ or ‘what action is best,’ regardless of the subject matter, is a moral decision (Sidgwick, 1874/1962). Therefore, decisions made in politics and economics can sometimes be understood as moral (Crespo,1998; Mill, 1863). Moral decisions have been found to be affected by either emotion (De La Viña et al., 2015; Conway et al., 2013; Haidt et al., 2001; Lee et al., 2015; Manfrinatti et al., 2013; Ugazio et al., 2012) or by deliberative cognition (Li et al., 2018; Paxton et al., 2012, Spears et al., 2021; Spears et al., 2018). Different political and economic orientations have also been associated with deliberative cognition (Deppe et al., 2015; Lane & Sulikowski, 2017; Yilmaz & Saribay, 2017a). Since political and economic beliefs are moral, and morality can be associated with deliberative cognition, here we attempt to analyze the moral underpinnings of political and economic beliefs, as well as their relations with deliberative abilities and thinking dispositions.

Previous studies investigating these connections, that will be reviewed later in this introduction, have used measurements for politics that have been mostly based on a singular self-reported orientation question. In this study we further include the SECS, as well as SFFM scales for political and economic attitudes, which will allow as to better disentangle not just liberal and conservative positions, but also to differentiate between social and economic conservatism (see below).

For deliberation, we also extend the measures. Besides using the CRT, we included the BNT, a deliberative numerical task that does not require cognitive reflection, and a verbal reasoning task, syllogisms, to estimate the belief bias (BBS). On the one hand, the addition of the

BNT may be of interest because the CRT has been found to reflect two separate measures of reflective abilities, where people can detect that there is a 'trick' in the question but lack the numeric ability to find the answer (Sinayev et al., 2015); the BNT will provided an index of the latter. On the other, the aim of the addition of the BBS was to include a deliberation measure that was not numeric based, as the BNT and CRT could be affected by math anxiety (Ashcraft, 2002; Ashcraft, & Ridley, 2005). Correct answers on both the BBS and BNT positively correlate with correct answers on the CRT (Toplak et al., 2014). While correct answers to both the BNT and BBS have both been found to be associated with utilitarian decisions on moral dilemmas (Baron et al., 2015; Spears et al., 2021), these measures of deliberation have not been used to study political and economic orientations.

Broadening the measurements of politics and economics will help create a clearer picture of connection between morality and politics. This is made even clearer by furthering the deliberative measures and their relationship to political and economic ideology. By understanding all of these connections could help improve communication, understanding, and compromise in politics and beyond.

### *1. Goals and Specific Hypotheses*

Thus, the objective of this study is to broaden the understanding and connections between morality, political and economic beliefs, and deliberation. Moral research has mostly been focused on hypothetical situations such as the trolley dilemma, and the next step is to broaden this connection to real world beliefs.

### *1.1. Relationship between political and economic beliefs, and moral judgments and decisions*

Using moral dilemmas, it has been shown an association between conservatism and deontological responses. In the present study, we aim to deepen this connection. For instance, Piazza & Sousa (2014) found that conservative participants, as measured with a single-question self-reported scale, thought that breaking a moral transgression (such as assisted suicide, lying, deception, betrayal, breaking the law, or treason) was less acceptable, even for the greater good. The rejection of these transgressions, regardless of the outcome by doing them, shows a conservative commitment to rules, or deontological thinking ('what you ought to do'). In addition, Antonenko Young et al.(2013) and (Chan (2019) found that participants who self-reported being more conservative than progressive politically made more deontological decisions when responding to moral dilemmas. Likewise, Lane et al. (2017) found that progressives took longer to respond to moral dilemmas under cognitive load, while conservatives' response times were not affected, giving further evidence to the hypothesis that conservatives emit immediate deontological responses to moral dilemmas. However, Chan (2019) found that while conservatism predicted deontological decisions, this was only true for people who were socially conservative participants (who believe that traditions and generational values should be maintained), while economically conservative people (who believe that inequality is necessary for society) did not show that pattern. This suggests that there is at least a crack in the conservative-deontological connection.

Therefore, in the present study, we use measures which allow us to disentangle social and economic conservatism (SECS and SFFM scales; see Materials) which may give us a better understanding of what aspects of political beliefs correlate with decisions on moral dilemmas or

not. We predict that measures of social conservatism will negatively correlate with utilitarian decisions on moral dilemmas (H3.a1), though it may not be true for economic conservatism (H3.a2).

*Conservatism and moral foundations*

A few researchers have already connected moral cognition with politics. Graham et al. (2009) found that there was a set of moral foundations, as measured by the Moral Foundations Questionnaire (MFQ; Care, Fairness, Loyalty, Purity, and Authority) that were connected with different political leanings. As a reminder, Care reflects how much a person tries to avoid harming others; Fairness represents the importance given to the equity of treatment of people; Loyalty refers to a person trusting those closest to them, and fears those who are not; Authority corresponds to the respect for hierarchies and those at the top, as long as those at the top provide protections and supplies; Purity reflects the feeling of disgust felt when another person's actions are not considered decent. Graham et al. (2009) found that the more liberal[17]/progressive a person reported being, the more they valued Care, and Fairness. Another study. by Crone and Laham (2015), found that Care (progressive foundation) and Purity (conservative foundation) predicted deontological responses to moral dilemmas. As Care and Fairness are highly endorsed by progressive participants, this brings into question the utilitarian-liberal connection, but supports the deontological-conservative connection. Comparing responses to these foundations with responses to moral dilemmas will help

[17] Liberal, in the United States, is a term referring to more socially liberal, where people should be allowed to do as they please as long as it does not harm others, and economically more left, where people believe the government and taxes ought to be used to benefit everyone in the nation. The American political party associated with liberalism (arguably) is the Democratic Party, and its Spanish equivalent would be The Spanish Socialist Workers Party (PSOE).

determine whether the general observed pattern can be replicated, and if it can be found in the Spanish population, under a rather different party-political system.

Therefore, we predict that the so-called conservative foundations (Loyalty, Purity, and Authority) will be negatively correlate with utilitarian decisions to moral dilemmas (H3b).

### *1.2. Politics, economics and deliberation*

Various studies have also looked at different political beliefs and their relations with deliberative cognition, finding that conservatism is often associated with intuitive cognition more than deliberative. For instance, the CRT has shown to be negatively correlated with self-reported conservatism (Deppe et al., 2015), as well as with the fact of being a Republican voter in the 2016 US Presidential elections (Pennycook & Rand, 2019). In addition, Chan (2019) found that self-reported intuitive cognitive styles positively correlated with conservatism. However, more economic conservative participants performed better on the CRT (Deppe et al., 2015; Pennycook et al., 2018), and reported using more deliberative cognition when making decision (Chan, 2019).

Many of these studies are based on a single-question measurement for political beliefs, and the CRT, a numerical task, to measure cognitive styles (Deppe et al., 2015; Pennycook et al., 2018). In the same vein, Chan (2019) had participants respond to questions on social and economic issues to determine their political stances, but measured deliberation not by abilities performing a task engaging deliberation, but through a self-reporting questionnaire on whether the participant makes decisions based on intuition or deliberation (Preference for Intuition or Deliberation Scale, PIDS; Betsch, 2004).

While all these studies add insight into the political and economic beliefs' connections with cognition, they do not go far enough on either measure of politics or deliberation. To better

understand the relationship between political/economic ideology and cognition, different, deeper, measures of politics, economics, and, particularly, deliberative cognition are needed. This study tries to bridge that gap in the literature.

Regarding moral foundations, Wright & Baril (2011) found that when participants complete a cognitively difficult task before responding to the MFQ, support for Purity, Loyalty, and Authority (considered as conservative foundations) decreased significantly for conservative participants, which implies that the foundations mainly supported by conservative participants take more effortful cognition. However, Van Berkel et al. (2015) found that cognitive load and induced low effort thought increased support for Authority, and decreased support for Fairness, respectively calling into question the findings from Wright et al. (2011). Regarding the more liberal foundations, Yilmaz & Saribay, 2017b found that support for Care and Fairness increased after responding to the CRT before answering the MFQ (compared to after), and that giving feedback on incorrect answers to the CRT also increased support for the same foundations, highlighting the liberal-deliberative connection.

Given this, the hypothesis is that conservative thinking, as measured by political orientation and both the SECS and the SFFM, will correlate negatively with performance on the deliberative tasks (H3.c), and that that correct answers on deliberation tasks will negatively correlate with Loyalty, Purity, and Authority (H3.d).

## 2. METHODS

### *2.1. Participants*

Participants (N = 198; 125 females; age, *M* = 30.01, *SD* = 11.62, range = 18 - 69), were recruited online; those doing their bachelor's in Psychology at the University of Granada were

given credits (n = 98) while the rest were not compensated. Participants had different education levels, though approximately 60% had at least a bachelor's degree ($n_{bachelor's}$ = 77; $n_{master's}$ = 32; $n_{doctorate}$ = 13), while the remaining participants had finished secondary education (n = 74), not finished secondary education (n = 1), or has never studied (n = 1). Almost all of the participants reported their 'mother language' being Spanish, except for Catalan (n = 3), Galician (n = 2), and one German speaking participant (n = 1).

### *2.2 Materials*

All materials were in Spanish, and presented online using Lime survey.

#### *2.2.1 Moral dilemmas*

There were five dilemmas that were taken from Moore et al. (2008) and Spears et al. (2018). Participants responded to a question similar to *'should you sacrifice this person in order to save yourself and five other people'* by using a six-point Likert scale anchored where 1 was completely against the sacrificing, and 6 was totally for sacrificing (see Bartels et al., 2011), in order to insure there were no neutral responses to the dilemmas. Responses between 1 – 3 were considered to support deontological decisions, while answers 4 – 6 were responses endorsing utilitarian decisions. The dilemmas chosen were *Crying baby, Submarine* (Spears et al., 2018), *Rescue 911, Bus plunge, and Cinder block* (Moore et al., 2008) which can be found in Appendix E.1.

#### *2.2.2 Deliberation Tasks (Cognitive Reflection Test (CRT), Berlin Numeracy Test (BNT), and Belief Bias syllogisms (BBS)*

All deliberation materials can be found in Appendix E.2.

*2.2.2.1. CRT*

To measure the reflective deliberation, we used a new set of CRT questions from Primi et al. (2015). This new CRT was also made up of three numeric questions, such as '*Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class?'* with the intuitive answer being 30 and the correct answer being 29. Correct answers were coded as 1, incorrect were coded 0, and the total scores were tallied an analyzed.

*2.2.2.2. BNT*

All four questions of the BNT (Cokely et al., 2012) were also used. One example question of this task is: '*imagine we are throwing a 5-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number?'* The scores were coded and tallied similarly to the CRT.

*2.2.2.3. Syllogisms task (Belief bias)*

We used eight syllogisms taken from De Neys and Van Gelder, (2009) that were separated into four different categories: Valid-believable (VB); Valid-unbelievable (VU); Invalid-believable (IB); and Invalid-unbelievable (IU). VB syllogisms are those whose conclusions are believable, and the conclusion logically follows from the premises. For example, '*all birds have wings; eagles are birds; therefore, eagles have feathers*. VU syllogisms are those whose conclusions are unbelievable but is still valid. For example, '*all mammals can walk; whales are mammals; therefore, whales can walk'*. IB syllogisms are ones whose conclusions are believable, but invalid. For example, *'all unemployed people are poor; Marta Gonzalez is not unemployed; therefore, Marta Gonzalez is not poor*. Finally, IU are syllogisms where the conclusion is unbelievable and not valid. For example, *'All pistols are dangerous; swords are dangerous; therefore, pistols are*

*swords*.' Participants responded wither 'yes' or 'no' to whether the conclusion logically followed from the premises, and were coded as 1 for correct answers, 0 for incorrect answers, with the total of correct answers tallied for analysis.

### *2.2.3 Political and economic orientation*

#### *2.2.3.1. Political orientation.*

Participants were asked '*when it comes to politics, do you generally consider yourself progressive, moderate, conservative, or another option'*, and self-reported their political orientation by means of a single-item 7-point Likert scale anchored from 1 'Very Progressive', to 7 'Very Conservative' (Inbar et al., 2009). In addition, they were asked *'generally, do you vote or identify with a particular political party'* and participants responded by writing their own responses. This scale can be found in Appendix E.3.1.

#### *2.2.3.2. Moral foundations*

The moral foundation questionnaire, MFQ (Haidt & Graham, 2007) measures five different moral foundations: Care, Fairness, Loyalty, Purity, and Authority. Participants were given sixteen scenarios and asked '*when you decide whether something is right or wrong, to what extent are the following considerations relevant to your thinking?'* before responding to a set of questions which were scored from 0 (not relevant) to 5 (extremely relevant). Another sixteen questions asked participants to what extent they agree or disagree to questions similar to *'Compassions to those who suffer is a virtue'*. Participants responded from 0 ('totally disagree') to 5 ('totally agree'). In total, there were thirty-two questions. Total scores were tallied among the different measures of Care, Fairness, Loyalty, Purity, and Authority and each measure was analyzed separately. This questionnaire can be found in Appendix E.3.2.

*2.2.3.3. Social and Economic Conservatism Scale (SECS; Everett, 2013)*

This is a 12-item questionnaire which distinguishes between social and economic conservatism. It was translated from English into Spanish by the author, a native English speaker, and reviewed by a native Spanish speaker. Participants are asked to rate how they felt about certain subjects stated in a single word or short phrase (e.g., abortion or social security) using a response scale between 0 (lowest degree of agreement) and 100 (highest degree of agreement). It measures conservative beliefs, though it consists of two sub-scales: SECS-Social (7 items; SECS-S) and SECS-Fiscal (5 items; SECS-F) measures. Some social issues are abortion, traditional values, or religion; while some fiscal questions are social security, fiscal responsibility, or business. This questionnaire can be found in Appendix E.3.3.

*2.2.3.4. Support for the Free Market System (SFFM) scale (Heath, & Gifford, 2006)*

This is a 6-item questionnaire about economic ideology. Participants were asked '*to what extent do you agree or disagree'* with questions such as: *an economic system based on free markets unrestrained by government interference automatically works best to meet human needs*. Participants responded to questions like these on a scale from 1 (totally disagree) to 4 (totally agree). This scale can be found in Appendix E.3.4.

***2.3 Procedure***

The experiment was performed online, using the platform Lime Survey provided by the University of Granada. After reading and accepting the informed consent, participants were sent to a page where they entered their personal information on socio-demographics, including age, gender, location, level of education, area of study and/or occupation. After being provided with basic instructions, participants first responded to the dilemmas in a random order. Following this,

participants filled in the different questionnaires starting with the MFQ, Self-reported political orientation item, SECS, SFFM, the BNT, CRT, and, finally, the syllogisms task[18].

## 3. RESULTS

### *3.1. Utilitarianism, Political and economic beliefs, and Moral Foundations*

First, we ran a series of preliminary analyses so check general patterns: link between deliberation and utilitarianism, relationships between the different measures of conservatism, as well as an assessment of the underlying two-factor structure in moral foundations (conservative foundations vs libera foundations). After that, we ran specific analyses to assess our particular hypotheses.

#### *3.1.1. Deliberation and Utilitarianism*

To check previous research on deliberative measures and utilitarianism, correlations between average utilitarian response to dilemmas and scores on the deliberative tasks were performed. Utilitarian decisions, as expected, positively correlated with correct answers on the CRT ($r$ = .16, $p$ = .013), the BNT ($r$ = .25, $p$ < .001), and the BBS ($r$ = .23, $p$ = .001).

#### *3.1.2. Conservatism measures*

To check whether the four measures of conservatism correlate between them, a Pearson's correlations was ran on scores on the single-item self-reported political orientation, the total scores

[18] Participants also responded to these questionnaires (climate change belief, confidence in Unions scales, fake news, SRP-III). Data not included in this study.

on SECS-Social (SECS-S), SECS-Fiscal (SECS-F), and SFFM. As expected, it was found that all four measures positively correlated between each other (Table 5.1).

**Table 5.1.** *Correlations (one tailed) between political orientation, Social and Economic Conservatism Scale (SECS S and SECS F), and Support for the Free Market (SFFM).*

**Pearson's Correlations**

| Variable | | Political Orientation | SECS Social | SECS Fiscal |
|---|---|---|---|---|
| SECS S | Pearson's r | **0.628*** ** | — | |
| | p-value | < .001 | — | |
| SECS F | Pearson's r | **0.445*** ** | **0.527*** ** | — |
| | p-value | < .001 | < .001 | — |
| SFFM | Pearson's r | **0.368*** ** | **0.323*** ** | **0.395*** ** |
| | p-value | < .001 | < .001 | < .001 |

* p < .05, ** p < .01, *** p < .001

*3.1.3. Moral Foundations: two factor structure*

Score on the five moral foundations, each made up of six question items, were factor analyzed using Principal Component Analysis (PCA) with Varimax (orthogonal) rotation. The analysis yielded two factors explaining a total of 77.201 % of the variance for the entire set of variables. Factor 1 (RC1) was labeled Conservative foundations (Authority, Loyalty, and Purity), as these three factors strongly correlated with conservative political ideology (see table 5.2). This factor explained 47.37 % of the variance. The second factor (RC2) was labeled the Progressive foundations (Care and Justice), as previous research has found that participants who rate being more progressive generally only support these two foundations (Graham et al., 2011). The variance explained by this factor was 29.831 %.

The communalities of the variables included are quite strong (Table 5.3), with all variables in each factor having at least 70% of variance in common with other variables in the analysis. Further, the KMO and Bartlett's Test of Sphericity both indicate that the set of variables are at least adequately related for factor analysis. Essentially, there are two clear factors of endorsements of moral foundations: either Conservative or Progressive. These two factors are also independent of one another (i.e., they are not correlated).

**Table 5.2. Component Loadings**

| | **RC1** | **RC2** |
|---|---|---|
| | *Conservative* | *Progressive* |
| CARE | | 0.856 |
| FAIRNESS | | 0.919 |
| LOYALTY | 0.853 | |
| AUTHORITY | 0.903 | |
| PURITY | 0.822 | |

*Note.* Applied rotation method is promax.

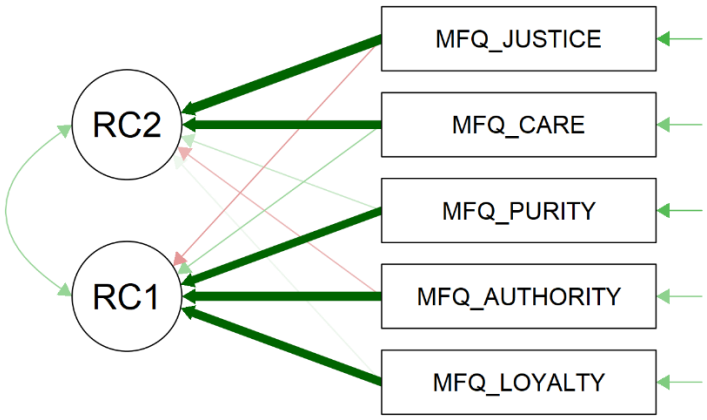


Figure 5.1 Principal Component Analysis: Path Diagram

*3.1.4. Moral foundations and conservatism measures*

To examine the relationship between moral foundations and conservatism, Pearson's correlations between the foundations and self-reported political orientation, the SECS, and SFFM scales (Table 5.3) was run. As a reminder to reads, Care and Justice were expected to correlate negatively with conservatism, while Loyalty, Authority, and Purity were thought to correlate positively.

**Table 5.3.** *Correlations between measures of political conservatism, and endorsements of moral foundations.*

**Pearson's Correlations**

| Variable | | Care | Fairness | Loyalty | Authority | Purity |
|---|---|---|---|---|---|---|
| Political Orientation | Pearson's r | -0.024 | **-0.300*** ** | **0.435*** ** | **0.495*** ** | **0.525*** ** |
| | p-value | 0.741 | < .001 | < .001 | < .001 | < .001 |
| SECS-S | Pearson's r | 0.024 | **-0.256*** ** | **0.563*** ** | **0.607*** ** | **0.544*** ** |
| | p-value | 0.734 | < .001 | < .001 | < .001 | < .001 |
| SECS-F | Pearson's r | -0.132 | **-0.275*** ** | **0.279*** ** | **0.347*** ** | **0.262*** ** |
| | p-value | 0.063 | < .001 | < .001 | < .001 | < .001 |
| SFFM | Pearson's r | **-0.273*** ** | **-0.422*** ** | **0.216** ** | **0.224** ** | **0.234*** ** |
| | p-value | < .001 | < .001 | 0.002 | 0.002 | < .001 |

* p < .05, ** p < .01, *** p < .001 (underlined, trend to significance)

*3.1.5 Utilitarianism and political and economic conservatism*

Correlations (two-tailed) between utilitarian response to each dilemma (*Crying baby, Submarine, Rescue 911, School bus, and Space station*), with the different measures of conservatism were ran. Scores on Self-reported political orientation, Social Conservatism and Support for the Free Market negatively correlated with some of the dilemmas (Table 5.4), Interestingly, Economic Conservatism positively correlated with responses to two of the dilemmas.

**Table 5.4.** *Correlations between measures of political conservatism, and utilitarian responses to the moral dilemmas*

| Variable | | Crying baby | Submarine | Rescue 911 | School bus | Space station |
|---|---|---|---|---|---|---|
| Political Orientation | Pearson's r | **-0.211** ** | **-0.172** * | -0.038 | **-0.141** * | -0.014 |
| | p-value | 0.003 | 0.015 | 0.599 | 0.047 | 0.845 |
| SECS Social | Pearson's r | **-0.159** * | -0.047 | -0.007 | -0.071 | 0.055 |
| | p-value | 0.025 | 0.510 | 0.925 | 0.320 | 0.439 |
| SECS Economic | Pearson's r | 0.012 | -0.086 | 0.082 | 0.126 | **0.166** * |
| | p-value | 0.866 | 0.231 | 0.251 | 0.077 | 0.020 |
| SFFM | Pearson's r | **-0.206** ** | **-0.144** * | -0.024 | -0.025 | -0.090 |
| | p-value | 0.004 | 0.044 | 0.737 | 0.728 | 0.205 |

* p < .05, ** p < .01, *** p < .001 (underlined, trend to significance)

*3.1.6 Utilitarianism and political and moral foundations*

In general, although scarce (Table 5.5) correlations were positive for liberal foundations (Care, Justice) and negative for Conservative ones (Purity).

**Table 5.5.** *Correlations between scores on moral foundations, and utilitarian responses to the moral dilemmas*

| Variable | | Crying baby | Submarine | Rescue 911 | School bus | Space station |
|---|---|---|---|---|---|---|
| CARE | Pearson's r | 0.042 | 0.139 | 0.064 | -0.052 | -0.071 |
| | p-value | 0.556 | 0.051 | 0.367 | 0.466 | 0.323 |
| FAIRNESS | Pearson's r | 0.130 | **0.205 **** | 0.082 | 0.053 | -0.001 |
| | p-value | 0.068 | 0.004 | 0.251 | 0.458 | 0.986 |
| LOYALTY | Pearson's r | -0.031 | 0.082 | 0.081 | -0.019 | 0.025 |
| | p-value | 0.668 | 0.252 | 0.256 | 0.795 | 0.729 |
| AUTHORITY | Pearson's r | -0.084 | 0.040 | 0.028 | -0.102 | 0.077 |
| | p-value | 0.238 | 0.581 | 0.697 | 0.153 | 0.283 |
| PURITY | Pearson's r | -0.135 | -0.035 | 0.002 | **-0.173 *** | -0.086 |
| | p-value | 0.058 | 0.620 | 0.977 | 0.015 | 0.228 |

* p < .05, ** p < .01, *** p < .001 (underlined, trend to significance)

***3.2 Deliberation and political and economic beliefs***

Similarly, measurements on responses to dilemmas, Pearson's correlations with measures of deliberation and political and economic beliefs were used.

*3.2.1 Deliberation and conservatism*

First, we correlated scores on the CRT, BNT, and BBS with those of the self-reported political orientation, the SECS-S, SECS-F and SFFM (Table 5.6) All measures of conservatism

correlated negatively with BBS, SFFM scores negatively correlated with that of CRT and BBS, while there as a trend toward significance in the correlations between BNT and both political orientation and SECS-S. Thus, generally, we found negative relationships between measures of deliberative thinking and conservatism, in this case either social or economic.

**Table 5.6.** *Correlations between deliberative measures (CRT, BNT, BBS) with political measures (Political Orientation, SECS, and SFFM).*

**Pearson's Correlations**

| Variable | | Political Orientation | SECS-S | SECS-F | SFFM |
|---|---|---|---|---|---|
| CRT | Pearson's r | -0.035 | -0.110 | 0.013 | **-0.172 *** |
| | p-value | 0.622 | 0.124 | 0.856 | 0.015 |
| BNT | Pearson's r | -0.136 | -0.128 | -0.027 | -0.102 |
| | p-value | 0.056 | 0.073 | 0.710 | 0.154 |
| BBS | Pearson's r | **-0.231 **** | **-0.268 ***** | **-0.148 *** | **-0.263 ***** |
| | p-value | 0.001 | < .001 | 0.038 | < .001 |

* p < .05, ** p < .01, *** p < .001 (underlined, trend to significance)

### *3.3 Deliberation and moral foundations*

As can be seen in Table 5.7, the conservative moral foundations showed negative correlations with measures of deliberative thinking, particularly Authority and Purity, and, remarkable, all of them with BBS.

**Table 5.7.** *Correlations between measures of deliberation (CRT, BNT, and BBS) with the five moral foundations (Care, Fairness, Loyalty, Authority, and Purity)*

**Pearson's Correlations**

| Variable | | Care | Fairness | Loyalty | Authority | Purity |
|---|---|---|---|---|---|---|
| CRT | Pearson's r | -0.073 | 0.079 | -0.056 | **-0.172 *** | **-0.227 **** |
| | p-value | 0.305 | 0.270 | 0.434 | 0.015 | 0.001 |
| BNT | Pearson's r | -0.070 | 0.079 | -0.070 | **-0.217 **** | **-0.240 ***** |
| | p-value | 0.330 | 0.268 | 0.326 | 0.002 | < .001 |
| BBS | Pearson's r | 0.003 | 0.117 | **-0.210 **** | **-0.305 ***** | **-0.314 ***** |
| | p-value | 0.967 | 0.102 | 0.003 | < .001 | < .001 |

* p < .05, ** p < .01, *** p < .001 (underlined, trend to significance)

## 4. DISCUSSION

In this study, we set out to broaden our understanding of the connections between utilitarian decisions and political and economic orientations, as well as to further test the connections between political orientation and deliberative abilities. This was done using two measures of morality (moral dilemmas and moral foundations), four measures of political and economic orientation (single-item political orientation, SECS, both social and economic, and SFFM), as well as three measures of deliberative thinking (CRT, BNT and BBS). With respect to moral foundations, we found evidence confirming the connection between the three so-called conservative moral foundations and conservatism (Graham et al., 2011) in our Spanish sample.

Regarding our particular hypotheses on utilitarianism and conservatism, we found evidence showing negative correlations between most measures of *conservatism* and utilitarian responses to moral dilemmas (supporting H3.a1), with the exception of the positive relationship in

the case of economic conservatism (thus confirming H3.a2). This is consistent with the results of the study by Chan (2019) who found that only social conservatism, and not fiscal conservatism, negatively correlated with utilitarianism.

Similarly, there was partial evidence on the negative connection between *conservative foundations* and utilitarian responses to moral dilemmas (H3.c). More strong evidence was found with respect to either the negative relationship between both *conservatism* (H3.c) or *conservative moral foundations* (H3.d) and measures of *deliberative thinking*, particularly in the case of the Belief Bias Syllogism (BBS) task, a measure of deliberative, as opposed to intuitive, reasoning. Thus, evidence of the relationship between utilitarianism and socio-economic political position or their underlying moral foundations was rather weak, while the negative relationship between socio-economic conservatism and deliberative cognition was strong in terms of syllogistic reasoning but weaker in reflective-numeric tasks.

Finding negative relationships between all measures of conservatism and deliberative thinking in the present study contradicts the finding of Chan (2019) that only social conservatives reported that they rely more on intuition. A key difference between the study reported here and Chan's is that standardized tasks were used in this study which measure deliberative abilities, while he used a self-reported measure. This is particularly important because, as Frederick (2005) found, participants who perform worse on the CRT are more confident in their answers compared to participants who answered more questions correctly, suggesting that people are not particularly good at assessing their own cognitive abilities.

Further, finding a negative correlation between correct answers on the CRT and SFFM, but not with the BNT, may add insight to the different types of cognition that is associated with

economic conservatism. Sinayev et al. (2014) found that responding correctly to the CRT is a two-step process: 1) a reflective step where a person recognizes that there is a trick or a lure that they must overcome, and once they overcome that lure 2) they need the numeric ability, or the ability to understand how numbers and proportions work together (Ghazal et al., 2014), to find the correct answer. To answer questions correctly on the BNT, a person only needs to rely on the later abilities and not on the former. Therefore, the negative correlation between the CRT and SFFM suggest that fiscal conservatives may be less reflective compared to more fiscally progressives.

However, to overcome problems with deliberation as measured by numeracy which may be affected by math anxiety (Morsanyi et al., 2014), we included Belief Bias Syllogisms, a verbal deliberative task without numeric load. Belief Bias Syllogisms has shown to positively correlate with the CRT (Baron et al.,2015), and analytical cognitive styles (Trippas et al., 2015). Taken into account our results, they suggest that there may be few differences between conservatives and progressives in numeric abilities which require reflection. However, when it comes to deliberative abilities as measured through logical syllogisms (reasoning) the picture is much clearer, regardless of whether they support fiscal conservative beliefs or social conservative beliefs: They perform worst.

The moral foundations of Loyalty, Purity, and Authority (conservative ones) correlated negatively with the three different measures of deliberation. This further contradicts the findings from Wright et al. (2011), who found that by exhausting cognitive resources before responding to the MFQ, decreased support for Loyalty, Purity, and Authority, suggesting that supporting these foundations use more cognitive resources (deliberative thinking). The data presented here come closer to supporting Yilmaz et al. (2015) work, where participants who performed the CRT before responding to the MFQ had increased endorsements of Care and Fairness with no effect on Loyalty,

Purity, and Authority. In our study, interestingly, Purity and Authority negatively correlated with correct answers on the CRT, BNT, as well as on the BBS, which implies that people who support these foundations may tend to be less reflective (CRT), less deliberative (BNT), and less able to overcome their own beliefs to understand logical conclusions (BBS).

*4.1. Conclusions and Limitations*

The current findings suggest that more conservative ideology, regardless of social or fiscal conservatism, are associated with poorer syllogistic reasoning; that is, they are worse at overcoming their own biases and finding logical conclusions from premises. Endorsement for the conservative moral foundations of Purity and Authority were also associated with poorer performance on syllogistic reasoning, but also on numeracy and numeric reflection. The foundation of Loyalty was also negatively associated with correct answers to syllogistic reasoning. This further suggests that the endorsements of Purity and Loyalty are based less on reflective, deliberative cognition, and logical cognition, and possibly more on intuition.

There are a few limitations to keep in mind when interpreting these findings. Firstly, and importantly, the number of self-reported conservative participants (n = 17) compared to liberal (n = 143) and moderates (n = 38) participants, was exceptionally low, which may weaken the relationships found in this study, even if they closely reflect the main patterns found in previous literature pointing out to a conservative-intuition link. Similarly, responses to different types of moral dilemmas might have different correlations with political orientations. Personal self-inevitable dilemmas were used in this study, and perhaps conclusions might be different for impersonal, ‘other’ dilemmas, and/or ‘inevitable’ dilemmas. Further, congruent and incongruent dilemmas could be used with process dissociation analysis to see if conservatives and/or

progressives support utilitarian decisions, or if they are simply non-deontological (Conway et al., 2013). Therefore, future studies should use a larger variety of dilemmas and measure them alongside more specific measures of political and economics. Another limitation is that the scales SECS and SFFM have not been validated in the Spanish population, therefore this is just a preliminary exploratory study. Finally, the simple correlational analyses do not go deep into the causal relationships between morality, politics and economics, and deliberation.

# Part IV: Discussion and Conclusion

# Chapter 6

# General Discussion, Conclusions, and Future perspectives

## 6.1. General Discussion

The main objective of this thesis was to explore the relationship between morality and deliberative cognition. To this end, the present work is structured into three sections. In two studies in Chapter 3 (Study 1, and 2), utilitarian decisions were increased by inducing deliberative cognition, using two different methods: numeric tasks with feedback, and difficult to read fonts; but were decreased by using cognitive control tasks, Stroop and Flanker. In Chapter 4, Study 4, the connection between utilitarian decisions, cognitive reflection, and psychopathy were examined in order to test if there was a relationship between deliberation and blunted social emotions in. In Chapter 5, the relationships between utilitarian decisions on moral dilemmas, political and economic conservatism, and deliberative cognition were estimated. Before discussing the theoretical and real-world implications, a brief summary of the findings is presented.

A. In the studies included in Chapter 3, the main objective was to induce deliberative cognition to increase utilitarian decisions. Three experiments were run to test this objective. The basic procedural details, predictions, findings, and conclusions of these studies are as follows.

1. Study 1 (*Can induced reflection affect moral decision-making?*) made use of two different numerical tasks and feedback before responding to dilemmas to induce utilitarian responses.

a) Participants either responded to the Cognitive Reflection Test (CRT) or the Berlin Numeracy Test (BNT) before responding to moral dilemmas.

b) Two groups received feedback and a second chance to respond (CRT_Fb, BNT_Fb), and two other groups did not (CRT, BNT).

c) It was predicted that participants who receive feedback will make more utilitarian decisions compared to participants who do not (H1.a).

d) It was further predicted that participants who receive feedback on the CRT will make more utilitarian decisions compared to participants who receive feedback on the BNT (H1.b)

e) It was found that participants who received feedback in the BNT made more utilitarian decisions compared to other groups, therefore partially confirming H1.a (but not supporting H1.b)

f) This shows that numeric tasks can induce utilitarian decisions but also that it is not reflection that necessarily induces these decisions

2. In Study 2, participants read dilemmas in both easy (fluent) and difficult (disfluent) to read fonts.

a) Participants read both personal and impersonal dilemmas.

b) Participants read dilemmas in easy and difficult to read fonts.

c) It was predicted that participants will make more utilitarian decisions after reading dilemmas in a disfluent font (H1.c).

d) It was further predicted that disfluency would affect responses to personal moral dilemmas more strongly compared to impersonal fonts (H1.d).

e) It was found that:

(1) There was an increased in endorsement of utilitarian decisions when dilemmas were written in a disfluent font and were presented in first order, partially confirming H1.c.

(2) There were no effects of fluency on type of dilemma, and therefore we were unable to confirm H1.d.

f) This shows that difficult to read fonts can induce utilitarian decisions on moral dilemmas, which may be due to deliberative cognition. However, as this was contingent upon the order, this mostly holds under some conditions.

3. In Study 3, participants responded to dilemmas after completing blocks of High Load and Low Load cognitive control tasks.

a) It was predicted that utilitarian decisions would increase after High Load blocks of cognitive control compared to responses after Low Load blocks of cognitive control.

b) There were no differences of responses on moral dilemmas regardless of whether they came after High or Low load blocks of the Stroop task.

c) Endorsement of utilitarian decisions decreased after High conflict blocks of the flanker task.

d) It is possible that High Load cognitive control tasks lead to cognitive fatigue, which decreases support for utilitarian decisions in moral dilemmas. While this does not confirm H1.e, the data might still show support for the deliberative-utilitarian connection.

B. In Chapter 4, the objective was to study the relationships between responses on moral dilemmas, psychopathic trait, and cognitive reflection (correct answers on the CRT).

1. It was predicted that correct answers on the CRT would positively correlate with utilitarian decisions (H2.a); that psychopathic trait would also correlate positively with utilitarian decision (H2.b); but that correct answers on the CRT and psychopathic trait would not correlate with each other (H2.c).
2. Endorsement of utilitarian decisions correlated positively with correct answers on the CRT, confirming H2.a.
3. Psychopathic trait score correlated with utilitarian decisions on moral dilemmas, also confirming H2.b.
4. There were no correlations found between correct answers on the CRT and total score on psychopathy or in the CA (callous affect) subscale, thus confirming H2.c.
5. These findings show that there may be two independent paths to utilitarianism: cold and calculating personality, or deliberative thinking disposition.

C. In Chapter 5, the main objective was to study the relationship between moral decisions, Moral Foundations, political and economic beliefs, and deliberative abilities. It was predicted that:

1. Endorsements of utilitarian decisions on moral dilemmas would negatively correlate with measures of social conservatism, but not fiscal conservatism (H3.a).
2. Endorsements of utilitarian decisions would negatively correlate with the so-called conservative Moral Foundations (Loyalty, Purity, and Authority) (H3.b).

3. Correct responses on deliberation tasks would negatively correlate with measures of conservatism (H3.c).
4. Correct responses on deliberation tasks would negatively correlate with so-called conservative Moral Foundations (H3.d).

It was found that:

1. Responses to some dilemmas negatively correlated with conservative measures of self-reported political orientation, SECS-S, and SFFM while responses to other dilemmas positively correlated with self-reported political orientation and SFFM. This partially confirms H3.a since the relationship between conservatism and utilitarian decisions was found for some dilemmas, but not others.
2. Responses to some moral dilemmas positively correlated with the endorsement of some liberal moral foundations, while a few conservative moral foundations negatively correlated with utilitarian decisions. All of this further partially confirms H3.b.
3. Measures of political and economic conservatism negatively correlated with correct responses on deliberative cognition tasks. Specifically, BBS negatively correlated with all measures of conservativism, showing that participants who endorse more conservative ideologies are worse at overcoming their own biases to come to logical conclusions. Further, high scores on the SFFM, where participants reported more support for the free market, negatively correlated with correct answers on the CRT. Therefore, H3.c was partially supported by the data.

4. Correct responses on the CRT, BNT, and BBS negatively correlated with endorsements of the moral foundations of Purity, and Authority, which are two conservative foundations. The endorsement of Loyalty also negatively correlated with correct answers on the BBS. This gives partial, but strong support to H3.d, since the majority of the conservative foundations negatively correlated with the majority of the measures of conservatism.

Following are theoretical and real-world implications derived from these results.

### 6.1.1 Theoretical Implications

The several studies presented in this doctoral thesis have contributed theoretically to the understanding of deliberation and moral decision making through 1) the induction of deliberative cognition; and 2) addressing the criticisms of the deliberative-utilitarian connection, such as emotional blunting.

#### *6.1.1.1 Utilitarianism and deliberative thinking*

The dual-process model of moral decision making (Greene et al., 2004; Greene et al., 2001) states that utilitarian decisions on moral dilemmas come from a slow, deliberative cognition, while deontological are fast, intuitive and emotional. The connection between negative emotion and moral choices is supported behaviorally by the fact that participants who make more utilitarian decisions feel more guilt and sadness afterwards (Choe et al., 2011). It appears that negative emotions have the ability to affect moral decisions in such a way as to push decisions towards more rule based moral norms. Though, it has also been shown that utilitarian decisions increased when counteracting negative emotions by positive emotions (Gawronski et al., 2018) or by

emotional reappraisal (Lee et al., 2015). These results further indicate that emotional responses to dilemmas are not static, and instead are dependent upon the type of emotion.

Conversely, as explained in detail in the Introduction, other studies have found connections between deliberative cognition and utilitarian decisions (Baron et al., 2015; Li et al., 2018; Paxton et al., 2012). These studies help to affirm that utilitarian decisions on moral dilemmas can come from System 2 cognition and suggest that deliberative cognition could be induced. Other research has shown decreased utilitarian decisions after completing cognitively difficult tasks, exhausting cognitive resources (Greene et al., 2008; Timmons et al., 2018; Témolière et al. 2012) suggesting that, when participants respond to dilemmas after having their cognitive resources over worked, they default to intuitive responses.

The data presented in this thesis (Chapter 3, Studies 1 and 2) have built upon and contributed to this understanding of moral cognition. Firstly, and importantly, this thesis has shown that deliberative cognition may, not only be induced through various tasks, but that this may lead to increased utilitarian decisions. This supports the dual-process theory of morality of the utilitarian-deliberative connection. System 2 cognition is deliberative (Evans, 2008; Evans & Stanovich, 2013; Evans, 2011), and a task such as the BNT is associated with deliberative cognition (Cokely et al., 2012; Ghazal et al., 2014). Therefore, by having participants think more about how to answer these questions correctly, and improving them, may have induced an activation of Type 2 cognition.

Understanding the subtle, but important, difference between deliberation and reflection will help understanding not only that deliberation may affect moral decisions more than reflection but suggest the relevance of distinguishing between them in future research. More studies using

other types of deliberative and reflective cognition, outside of numeric tasks, could help sparse out this question. Future research might look at reflective word problems (Sirota et al., 2020), for example, with other word tasks of deliberative cognition.

What's more, this data may question a strong interpretation of the System 2- Utilitarian decisions bond, because it questions the reflective-utilitarian connection. That is to say, cognitive reflection may not be a mode of System 2 cognition that is associated with utilitarian decisions. That is not to say that reflection is not part of System 2 cognition, just not necessarily what affects utilitarian decisions. In Chapter 3, Study 1, Feedback after the CRT did not increase Utilitarian decisions and Feedback after the BNT did. While reflection was associated with Utilitarian decisions (Baron et al., 2015; Paxton et al. 2015) this may not support the relationship between reflection and utilitarian decisions since, to answer questions correctly on the CRT, a person is first reflective and then deliberative (Sinayev et al., 2015), therefore any association between correct responses on utilitarian decisions and correct answers on the CRT may be due to the deliberative aspect and not the reflective aspect of System 2 cognition. In the same vein, Byrd, and Conway (2019), found that performance on both CRT and BNT (numeric tasks) were associated with endorsements of utilitarianism, while performance on BBS (deductive reasoning task) was associated with endorsements on both utilitarian and deontological decisions. The authors proposed that this pattern of results would support a weaker version of the dual-process theory, where certain types of deliberative abilities would to be clearly connected to utilitarianism, but not others (like reflection).

Building on the data from Chapter 3, Study 1, disfluent fonts (Chapter 3, Study 2), which were controversially found to increase correct answers on the CRT (Alter et al., 2007; though not replicated, Thompson et al., 2013), lowered over-estimations of causal relationships in non-

contingent events (Díaz-Lago & Matute, 2019), and reduce acceptance of conspiracy theories (Swami et al., 2014), were found here to presumably make participants think more analytically, thus increasing utilitarian decisions. Once again, this added credence to the utilitarian-deliberation connection.

Finally, in Chapter 3, Study 3, though the intention was to induce deliberative cognition presenting participants with tasks prompting high-load cognitive control—similar to dot matrix tasks and letter tasks (Timmons et al., 2018)—in our case a reduction in the endorsement of utilitarian decisions were found in comparison with the low load blocks. As more research finds the utilitarian-deliberative connection (Li et al., 2018; Patil et al., 2021; Spears et al., 2018, 2021), it is just as important to understand what features of a task can reduce deliberative thinking, and thus reduce utilitarian decisions. Altogether, this helps our understanding of cognitive abilities, moral cognition, and also decision-making. That is, being able to induce deliberative cognition or inhibit it can help not only research in the moral domain, but also in decisions-making research, and research on dual-process theories.

A more difficult question to answer is why tasks such as the CRT, BNT or manipulations of disfluency increase utilitarian decisions, but other tasks such as Stroop or flanker did not. The Stroop task resulted in being more difficult than the flanker task. The reasons why high cognitive load reduces utilitarian decisions and why induced deliberation on the BNT and disfluency increase them can only be speculated upon. It could be that after completing a difficult cognitive control tasks, participants take a cognitive ‘break’ from effort by reading and responding to a dilemma before focusing on responding to the cognitive control tasks correctly. This may follow Lavie et al.’s (2014) load theory, where high concentration on one task can negatively affect the attention necessary for the following task, particularly when there’s dual-task switching (Lavie et

al., 2004). Conversely, working through difficult math problems as presented in Chapter 3, Study 1, may engage a cognitive mode which carries over sequential tasks, such as reading dilemmas. However, deliberation is not the only factor associate to utilitarian choices in moral dilemmas.

*6.1.1.2 Responses to criticisms of the Utilitarian-deliberative Connection.*

Some research has questioned the utilitarian-deliberative connection of the dual-process of morality. Bartels et al. (2011) argued that utilitarian decisions come from reduced emotions instead of increased deliberation, while Kahane et al. (2012) used dilemmas where the intuitive response is the utilitarian one. The research presented in this thesis may call into question the findings which are detailed below.

*6.1.1.3 Emotional Blunting or deliberation.*

Other pieces of research built on an alternative explanation: Utilitarian decisions come from a cold, calculating, emotionally blunted cognition instead of deliberative thinking (e.g., Bartels et al., 2006; Hayakawa et al., 2017; Koenigs et al., 2007; Lee et al., 2015). However, Conway et al., (2018) found that, psychopaths did not in fact endorse utilitarian decision, rather they endorsed harm, which weakens the claim of emotionally blunted utilitarianism. Adding to this, it has been reported in this thesis that correct answers on the CRT and Psychopathy (as measured by the SRP-III; Paulhus & Kelly, 2002) may be two different paths towards utilitarian decisions. Therefore, while blunted emotions may lead to more endorsements of utilitarian decisions, it does not mean that they are endorsing the utilitarian decisions because it is morally better (Conway et al., 2018), and it does not mean that these blunted emotions affect deliberative cognition. Importantly, it does not mean that when a person deliberates, their emotions are blunted. That is, deliberation may be separate from and not including a cold-calculating cognition. In fact,

in this thesis, it was shown that people have psychopathic traits (as measured by the SRP-III) do not answer more questions correctly on the CRT. Essentially, this shows that while reduced emotions and deliberative cognition are both associated with utilitarian decisions on moral dilemmas, they are not necessarily linked. That is, participants were either deliberative or cold and calculating, but not both.

#### *6.1.1.4 Morality: Moral foundations and utilitarianism*

Broadening the research on morality and moral theories, this thesis compared the endorsements of the different Moral Foundations and responses to moral dilemmas. Crone et al. (2015) consider that the foundations of Caring and Purity should negatively predict utilitarian decisions on moral dilemmas because the act of sacrificing one person to save five is neither a moral act of looking after or caring for another, nor is it pure since it does not show concern for life (Piazza et al., 2013). When looking at correlations with specific dilemmas, there were positive correlations between Care and some dilemmas; Justice with two dilemmas, while both two dilemmas negatively correlated with Purity. Further, contrary to previous research (Crone et al. 2015), Care and Justice both positively correlated with some Utilitarian decisions, though Purity did negatively correlate on some utilitarian decisions. This attempt to find a connection between the Moral Foundations Theory and the Dual-Process theory of Morality may be dependent upon the Foundation and the specific dilemma. Therefore, the connection between the two moral theories is tenuous.

### 6.1.2 Real-World Implications

It is one thing to study how deliberation is associated with utilitarian choices on moral dilemmas but understanding how deliberation is associated with real-world morality is also relevant. In accomplishing this, some researchers have used moral problems in scenarios including

self-driving cars (Bonnefon et al., 2015;Kallioinen et al., 2019), where people have been asked to decide whether a self-driving car should crash into a group of pedestrians in a crossing to save the driver, or if the car should swerve to avoid crashing into the pedestrians even if this means the driver will die. Others have looked at the political connections between moral foundations and politics (Graham et al., 2009; Graham et al., 2011). This thesis includes a focus on the later real-world scenario.

The connection between politics and morality can be quite strong. Philosophically, John Stuart Mill, an early utilitarian philosopher, spoke about the connections between utilitarianism and justice and laws (Mill, 1863) explaining that fines and punishments are forms of normative morality. Examining a few items on the Social and Economic Conservatism Scale (SECS; Everett, 2013), such as agreement with abortion, religion, and traditional values, the morality of such questions can be considered both political and moral. In fact, the trolley dilemma originated in an article about abortion (Foot, 1967) in connection with the Doctrine of Double Effect (Aquinas, 1274) which argues that even if harm occurs, as long as the intention of the action was not to harm that action is still morally permissible. The point being that, in certain situations, as long as the intention to abort was not intended, but an unfortunate consequence of some other action (such as in an attempt to save the mother's life in terms of abortion), then that action is morally permissible; the same logic could apply to the changing the train track to save others in the trolley dilemma. Support for or against abortion is a moral issue that is also a political issue as well (McKeegan, 1993).

Similarly, economic beliefs may also be considered moral. For instance, Crespo (1998), argued that economics is a moral science and pointed out that several economists, such as Keynes, but also conservative economists like Roepke and Robbins, have supported or conceded this point.

Still, real-world examples of economic practices with moral consequences such as the use of child labor (Marx, & Engels, 1867) to increase profits can still be found today in such things as Nestlé's cultivation of cocoa (Clarke, 2015). Or considering that nearly 26,000 US citizens die a year due to a lack of health Insurance (Tanne & Dyer, 2008), when it is argued that the free market ought to provide the best healthcare (Coburn, 2008; Kristol, 1994). This thesis takes the view that these are economic decisions but also, moral problems, which are often defended by the DDE since the intention is not to harm children or the uninsured rather to maximize profits. Thus, it is important to understand the cognitive underpinnings of these moral-economic choices such as letting children be harmed whilst working, particularly in unsafe conditions. That is, are those who support such free market beliefs more or less reflective, deliberative, or logical in the way they make judgements of economic issues? The data presented here might suggest that those ideologies which support these conditions for profit may not be associated with the most logical cognition.

To this end, the findings that political conservatism was negatively associated with correct answers to Belief Bias Syllogisms makes us wonder whether these economic decisions or ideologies are made without bias, or logically. In order to answer syllogisms correctly, participants need to overcome their own beliefs about the believability of the conclusion and then analyze whether the conclusion logically derives from the premises. However, participants that score high on conservative measures, both socially and fiscally, tend to do worse on answering these questions correctly. This may suggest that a person who believes that the free market will meet human's need, that this is more important than the environment (Heath et al., 2006), and against things such as social security, will do worse in overcoming their own biases to find logical conclusions.

These issues are not only contemporary problems but may have devastating effects like that commented before about 26,000 people dying a year from a lack of health care, as well as

including the looming environmental disaster (according to the World Health Organization). Also, women's rights to their bodies have recently been threatened, as countries like Poland have recently outlawed abortion, and the specter of a similar decision in the United States with the recent appointments of an anti-abortion Supreme Court Justices who can and may soon decide on such issues. While this thesis cannot present any solutions to these particular problems either, the understanding of the cognition behind such beliefs can help work towards an understanding how discussions on such topics may need to be centered in order to work towards future solutions. If biases towards one's beliefs can affect reasoning in a logically standardized scenario like syllogisms (i.e., the acceptance or negation of a valid conclusion which follows from premises), it is possible that this same effect can be more pronounced when it comes to not clear-cut logical problems in the real world. However, our data were correlational, thus precluding inferring conclusions about causal links between being conservative and draw incorrect conclusions from premises. However, it may be an indicator that, in political conversations and debates, logical reasoning may not be the strongest persuasive tool to use.

In his book *The Righteous Mind* (Haidt, 2012), Haidt explains that understanding the moral foundations of politics can be used as a means to communicate with people with other political orientations. For instance, since conservatives tend to value foundations such as Loyalty, Purity, and Authority (Graham et al., 2011), then appealing to these foundations may be more convincing than arguing from a point of Caring or Fairness. For example, when talking about traditional marriage, a social conservative question from the SECS (Everett, 2013), a person may have better luck talking about how gay marriage is a way for people to be Loyal, as well as Caring. This may not change minds, but it is appealing to their morality which might be more effective than trying to appeal to logic. This is ever more important since people who endorse Loyalty, Purity, and

Authority are also worse at overcoming their own beliefs to follow premises to logical conclusions. Essentially, logic and reasoning may not be the most effective way of conversing about politics in order to improve understanding, or to convince other people. In fact, recent research has shown that by non-judgmentally engaging in conversations with a person and eliciting stories or explanations about their experiences that have shaped their beliefs can do more to change opinions and is more effective than giving logical arguments (Kalla & Broockman, 2020).

**6.2 Conclusions:**

The conclusions of this thesis are as follows:

1. Participants who answered more questions correctly on tasks which measure deliberative abilities made more utilitarian decisions. This was found in participants who answered more questions correctly on both the Cognitive Reflection Test (Frederick, 2005) and Berlin Numeracy Test (Cokely et al., 2012), indicating that deliberation, but not necessarily reflective cognitive abilities, is associated with utilitarian decisions. This adds evidence for utilitarian-deliberative connection, supporting the dual-process model of morality (Greene et al., 20014; Greene et al., 2001), at least partially (Byrd et al., 2019).

2. Utilitarian decisions can be induced by having participants reconsider their wrong answers and given extra information to consider on the problem. Specifically, when participants are made to reconsider their wrong answers on the BNT (Cokely et al., 2012) they subsequently make more utilitarian decisions compared to participants who received feedback on

their incorrect answers on the CRT. This goes a step further in not just supporting the utilitarian-deliberative connection in the dual process model of morality: it also shows that deliberation can be induced and thus increases utilitarian decisions.

3. Difficult to read (disfluent) fonts compared to easy-to-read fonts increased utilitarian decisions (at least when the disfluent block is presented first), presumably because disfluency induces deliberative thinking.

4. Participants' responses to moral dilemmas were less utilitarian after completing a block of high load cognitive control task. However, this may be dependent on the type of task since this effect was found with a Flanker task, but not in the case of a color Stroop task, regardless of whether it was a high or low load cognitive control condition. While it was predicted that high-load cognitive control would increase utilitarian decisions, this data does not contradict the dual process model of morality. Previous research has shown that cognitive fatigue can reduce deliberative thinking and thus reducing utilitarian decisions (De Neys et al., 2017; Timmons et al., 2018) and that is what our data also suggest.

5. Utilitarian decisions correlated with correct answers on the CRT, but also on cold and calculating antisocial traits on the Self-Reported Psychopathy Scale (Importantly, correct answers on the CRT did not correlate with Callous Affect, a psychopathic trait, or psychopathy as a whole, indicating two independent paths to utilitarianism. Thus, utilitarian decisions can be

either cold and calculating or deliberative, but that deliberative cognition is not necessarily due to a cold and calculating personality.

6. When looking at dilemmas separately, there were negative correlations between utilitarian responses and Social Conservativism, and Support for the Free Market (SFFM); but positive correlations with Fiscal Conservatism. Likewise, when the Moral Foundations were compared with dilemmas separately, there were positive correlations with Care and Justice Foundation and negative correlations with Purity, suggesting a possible relationship between Foundations and utilitarianism, which might be dilemma dependent.

7. In general, the results of this thesis did not show negative correlations between the CRT score and any measurement of political orientation except for SFFM, yet Social Conservatism negatively correlated with score on the BN. However, all measures of conservatism negatively correlated with correct answers on Belief Bias Syllogisms. This means that more conservative participants may not be as capable of overcoming their own biases to find logical conclusions, though more conservative participants are necessary to fully understand these relationships.

8. Some Moral Foundations were also found to negatively correlate with deliberative cognition. The results in this thesis showed a similar pattern that was also found in previous literature, with negative correlations between Purity and Authority with the CRT, but not Loyalty. This same

pattern was found with the Berlin Numeracy Tes.t Additionally Loyalty, Purity, and Authority all negatively correlated with correct answers on Belief Bias Syllogisms. This shows how conservative foundations are related to unreflective, non-deliberative, and illogical cognition.

# References

Adolphs, R., Tranel, D., & Damasio, H. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature, 372*, 669-672 4.

Aguilar, P., Brussino, S., & Fernández-Dols, J.-M. (2013). Psychological distance increases uncompromising consequentialism. *Journal of Experimental Social Psychology, 49(3)*, 449–452. https://doi.org/10.1016/j.jesp.2013.01.002

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136(4)*, 569–576. https://doi.org/10.1037/0096-3445.136.4.569

Anscombe, G. E. M. (1958). Modern Moral Philosophy. *Philosophy, 33(124)*, 1–19. https://doi.org/10.1017/S0031819100037943

Antonenko Young, O., Willer, R., & Keltner, D. (2013). "Thou shalt not kill": Religious fundamentalism, conservatism, and rule-based moral processing. *Psychology of Religion and Spirituality, 5(2)*, 110–115. https://doi.org/10.1037/a0032262

Ashcraft, M. H. (2002). Math Anxiety: Personal, Educational, and Cognitive Consequences. *Current Directions in Psychological Science, 11(5)*, 181–185. https://doi.org/10.1111/1467-8721.00196

Ashcraft, M. H. (2005). *Math anxiety and its cognitive consequences: A tutorial review.* In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (P. 315 – 327). Psychology Press.

Aquinas, T. S. (1274). *The "Summa Theologica" of St. Thomas Aquinas. London:* Burns, Oates & Washburne.

Baron, J., & Gürçay, B. (2017). A meta-analysis of response-time tests of the sequential two-systems model of moral judgment. *Memory & Cognition, 45(4)*, 566–575. https://doi.org/10.3758/s13421-016-0686-8

Baron, J., Scott, S., Fincher, K., & Emlen Metz, S. (2015). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition, 4(3)*, 265–284. https://doi.org/10.1016/j.jarmac.2014.09.003

Barrash, J., Tranel, D., & Anderson, S. W. (2000). Acquired Personality Disturbances Associated With Bilateral Damage to the Ventromedial Prefrontal Region. *Developmental Neuropsychology, 18(3)*, 355–381. https://doi.org/10.1207/S1532694205Barrash

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition, 121(1)*, 154–161. https://doi.org/10.1016/j.cognition.2011.05.010

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting External Validity: Concerns about Trolley Problems and Other Sacrificial Dilemmas in Moral Psychology: External Validity in Moral Psychology. *Social and Personality Psychology Compass, 8(9)*, 536–554. https://doi.org/10.1111/spc3.12131

Bègue, L., Subra, B., Avers, P., Muller, D., Bricout, V., & Zorman, M. (2009). The message, not the bottle: Extrapharmacological effects of alcohol on aggression. *Journal of Experimental Social Psychology, 45*, 137-142. https://doi.org/10.1016/j.jesp.2008.07.018

Bentham, J. (1789). *An introduction to the principles of morals and legislation. Printed in the year 1780, and now first published. By Jeremy Bentham*. Printed for T. Payne, and Son.

Betsch, C. (2004). Präferenz für Intuition und Deliberation (PID). *Zeitschrift für Differentielle und Diagnostische Psychologie, 25(4)*, 179–197. https://doi.org/10.1024/0170-1789.25.4.179

Bialek, M., De Neys, W. (2017). Dual process and moral conflict: evidence for deontological reasoner' intuitive utilitarian sensitivity. *Judgment and Decision Making, 12(2)*, 148-167.

Bialek, M., & Terbeck, S. (2016). Can cognitive psychological research on reasoning enhance the discussion around moral judgments? *Cognitive Processing, 17(3)*, 329–335. https://doi.org/10.1007/s10339-016-0760-y

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science, 352(6293)*, 1573–1576. https://doi.org/10.1126/science.aaf2654

Bornstein, R.F., & D'Agostino, P.R. (1992). Stimulus recognition and the mere exposure effect. *Journal of Personality and Social Psychology, 63,* 545-552. Doi:10.1037/0022-3514.63.4.545.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108(3)*, 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Byrd, N., & Conway, P. (2019). Not all who ponder count costs: arithmetic reflection predicts utilitarian tendencies, but logical reflection predicts both deontological and utilitarian tendencies. *Cognition, 192*, 1-19. http://doi.org/10.1016/j.cognition.2019.06.007

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42(1),* 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119(2)*, 197–253. https://doi.org/10.1037/0033-2909.119.2.197

Carmona-Perera, M., Caracuel, A., Verdejo-García, A., & Perez-García, M. (2013). Adaptación a la población española de un instrumento de juicios morales: La batería de dilemas morales de Moore. *Anales de Psicología, 29(3)*, 827–835. https://doi.org/10.6018/analesps.29.3.137801

Carney, D. R., & Mason, M. F. (2010). Decision making and testosterone: When the ends justify the means. *Journal of Experimental Social Psychology, 46(4)*, 668–671. https://doi.org/10.1016/j.jesp.2010.02.003

Chan, E. Y. (2019). Social (not fiscal) conservatism predicts deontological ethics. *Acta Psychologica, 198:102867*. https://doi.org/10.1016/j.actpsy.2019.102867

Choe, S. Y., & Min, K.-H. (2011). Who makes utilitarian judgments? The influences of emotions on utilitarian judgments. *Judgment and Decision Making, 6(7)*, 13.

Ciaramelli, E., Muccioli, M., Làdavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience, 2(2)*, 84–92. https://doi.org/10.1093/scan/nsm001

Clifford, S., Iyengar, V., Cabeza, R., & Sinnott-Armstrong, W. (2015). Moral foundations vignettes: A standardized stimulus database of scenarios based on moral foundations theory. *Behavior Research Methods, 47(4)*, 1178–1198. https://doi.org/10.3758/s13428-014-0551-2

Coburn, T. (2008, March 10). Competition Solves Health Care. The New York Sun. https://www.nysun.com/opinion/competition-solves-health-care/72610/

Cokely, E. T., Galesic, M., Schulz, E., Garcia-Retamero, R., & Ghazal, S. (2012). Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making, 7(1)*, 23.

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making, 4(1)*, 14.

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology, 104(2)*, 216–235. https://doi.org/10.1037/a0031021

Conway, P., Goldstein-Greenwood, J., Polacek, D., & Greene, J. D. (2018). Sacrificial utilitarian judgments do reflect concern for the greater good: Clarification via process dissociation and the judgments of philosophers. *Cognition, 179*, 241–265. https://doi.org/10.1016/j.cognition.2018.04.018

Cooper, M. J., & Pullig, C. (2013). I'm Number One! Does Narcissism Impair Ethical Judgment Even for the Highly Religious? *Journal of Business Ethics, 112(1)*, 167–176. https://doi.org/10.1007/s10551-012-1239-0

Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., & Keysar, B. (2014). Your Morals Depend on Language. *PLoS ONE, 9(4)*, e94842. https://doi.org/10.1371/journal.pone.0094842

Costa, A., Vives, M., & Corey, J. D. (2017). On Language Processing Shaping Decision Making. *Current Directions in Psychological Science, 26(2)*, 146–151. https://doi.org/10.1177/0963721416680263

Crespo, R. F. (1998). Controversy: Is Economics a Moral Science? *Journal of Markets & Morality, 2*, 220 - 225.

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences, 107(40)*, 17433–17438. https://doi.org/10.1073/pnas.1009396107

Crone, D. L., & Laham, S. M. (2015). Multiple moral foundations predict responses to sacrificial dilemmas. *Personality and Individual Differences, 85*, 60–65. https://doi.org/10.1016/j.paid.2015.04.041

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44(1)*, 113–126. https://doi.org/10.1037/0022-3514.44.1.113

de la Viña, L., Garcia-Burgos, D., Okan, Y., Cándido, A., & González, F. (2015). Disentangling the Effect of Valence and Arousal on Judgments Concerning Moral Transgressions. *The Spanish Journal of Psychology, 18*, E61. https://doi.org/10.1017/sjp.2015.66

De Neys, W. (2012). Bias and Conflict: A Case for Logical Intuitions. *Perspectives on Psychological Science, 7(1)*, 28–38. https://doi.org/10.1177/1745691611429354

De Neys, W., & Van Gelder, E. (2009). Logic and belief across the lifespan: The rise and fall of belief inhibition during syllogistic reasoning. *Developmental Science, 12(1)*, 123–130. https://doi.org/10.1111/j.1467-7687.2008.00746.x

Deppe, K. D., Gonzalez, F. J., Neiman, J. L., Jacobs, C., Pahlke, J., Smith, K. B., & Hibbing, J. R. (2015). Reflective liberals and intuitive conservatives: A look at the Cognitive Reflection Test and ideology. *Judgment and Decision Making, 10(4)*, 19.

Díaz-Lago, M., & Matute, H. (2019). A hard to read font reduces the causality bias. *Judgment and Decision Making, 14*, 547-554.

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the ( ): Effects of disfluency on educational outcomes. *Cognition, 118(1)*, 111–115. https://doi.org/10.1016/j.cognition.2010.09.012

Duke, A. A., & Bègue, L. (2015). The drunk utilitarian: Blood alcohol concentration predicts utilitarian responses in moral dilemmas. *Cognition, 134*, 121–127. https://doi.org/10.1016/j.cognition.2014.09.006

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16(1)*, 143–149. https://doi.org/10.3758/BF03203267

Erlandsson, A., Nilsson, A., Tinghög, G., & Västfjäll, D. (2018). Bullshit-sensitivity predicts prosocial behavior. *PLOS ONE, 13(7)*, e0201474. https://doi.org/10.1371/journal.pone.0201474

Eskine, K. J., Kacinik, N. A., & Prinz, J. J. (2011). A Bad Taste in the Mouth: Gustatory Disgust Influences Moral Judgment. *Psychological Science, 22(3)*, 295–299. https://doi.org/10.1177/0956797611398497

Eslinger, P. J., & Damasio, A. R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology, 35(12)*, 1731–1731. https://doi.org/10.1212/WNL.35.12.1731

Evans, J. St. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning, 13(4)*, 321–339. https://doi.org/10.1080/13546780601008825

Evans, J. St. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology, 59(1)*, 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J. St. B. T. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review, 31(2–3)*, 86–102. https://doi.org/10.1016/j.dr.2011.07.007

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science, 8(3)*, 223–241. https://doi.org/10.1177/1745691612460685

Everett, J. A. (2013). The 12 item social and economic conservatism scale (SECS). *Plos One, 8(12)*: e82131. Doi: 10.1371/journal.pone.0082131.

Foot, P. (1978). The Doctrine of Double Effect. *A Companion to the Philosophy of Action, (5)*, 324–330.

Franklin, S. A., McNally, R. J., & Riemann, B. C. (2009). Moral reasoning in obsessive-compulsive disorder. *Journal of Anxiety Disorders, 23(5)*, 575–577. https://doi.org/10.1016/j.janxdis.2008.11.005

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives, 19(4)*, 25–42. https://doi.org/10.1257/089533005775196732

Gawronski, B., Conway, P., Armstrong, J., Friesdorf, R., & Hütter, M. (2018). Effects of incidental emotions on moral dilemma judgments: An analysis using the CNI model. *Emotion, 18(7)*, 989–1008. https://doi.org/10.1037/emo0000399

Ghazal, S., Cokely, E. T., & Garcia-Retamero, R. (2014). Predicting biases in very highly educated samples: Numeracy and metacognition. *Judgment and Decision Making, 9(1)*, 20.

Gigerenzer, G. (2010). Moral Satisficing: Rethinking Moral Behavior as Bounded Rationality. *Topics in Cognitive Science, 2(3)*, 528–554. https://doi.org/10.1111/j.1756-8765.2010.01094.x

Gold, N., Pulford, B. D., & Colman, A. M. (2013). Your money or your life: comparing judgments in trolley problems involving economic and emotional harms, injury and death. *Economics and Philosophy, 29(2)*, 213–233. https://doi.org/10.1017/S0266267113000205

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology, 96(5)*, 1029–1046. https://doi.org/10.1037/a0015141

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293(5537)*, 2105–2108. https://doi.org/10.1126/science.1062872

Greene, Joshua D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition, 107(3)*, 1144–1154. https://doi.org/10.1016/j.cognition.2007.11.004

Greene, Joshua D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The Neural Bases of Cognitive Conflict and Control in Moral Judgment. *Neuron, 44(2)*, 389–400. https://doi.org/10.1016/j.neuron.2004.09.027

Greene, Joshua David. (2014). Moral tribes: Emotion, reason, and the gap between us and them. Penguin Press.

Gürçay, B., & Baron, J. (2017). Challenges for the sequential two-system model of moral judgement. *Thinking & Reasoning, 23(1)*, 49–80. https://doi.org/10.1080/13546783.2016.1216011

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108(4)*, 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Haidt, J., & Graham, J. (2007). When Morality Opposes Justice: Conservatives Have Moral Intuitions that Liberals may not Recognize. *Social Justice Research, 20(1),* 98–116. https://doi.org/10.1007/s11211-007-0034-z

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus, 133(4),* 55–66. https://doi.org/10.1162/0011526042365555

Haidt, J., McCauley, C., & Rozin, P. (1994). Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Personality and Individual Differences, 16(5),* 701–713. https://doi.org/10.1016/0191-8869(94)90212-7

Handley, S. J., & Trippas, D. (2015). Dual Processes and the Interplay between Knowledge and Structure: A New Parallel Processing Model. In Psychology of Learning and Motivation (Vol. 62, pp. 33–58). Elsevier. https://doi.org/10.1016/bs.plm.2014.09.002

Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., & Keysar, B. (2017). Thinking More or Feeling Less? Explaining the Foreign-Language Effect on Moral Judgment. *Psychological Science, 28(10),* 1387–1397. https://doi.org/10.1177/0956797617720944

Heath, Y., & Gifford, R. (2006). Free-market ideology and environmental degradation: the case of belief in global climate change. *Environment and Behavior, 38(1),* 48-71. Doi:10.1177/001391650277998

Horberg, E. J., Oveis, C., Keltner, D., & Cohen, A. B. (2009). Disgust and the moralization of purity. *Journal of Personality and Social Psychology, 97(6),* 963–976. https://doi.org/10.1037/a0017423

Hume, D. (1777). *An Enquiry Concerning Human Understanding*. A. Millar.

Inbar, Y., Pizarro, D. A., & Bloom, P. (2009). Conservatives are more easily disgusted than liberals. *Cognition & Emotion, 23(4),* 714–725. https://doi.org/10.1080/02699930802110007

Johnson, E. J., & Goldstein, D. (2003). Do Defaults Save Lives? *Science, 302(5649),* 1338–1339. https://doi.org/10.1126/science.1091721

Kahane, G., Everett, J. A. C., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review, 125(2),* 131–164. https://doi.org/10.1037/rev0000093

Kahane, G., Everett, J. A. C., Earp, B. D., Farias, M., & Savulescu, J. (2015). 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition, 134,* 193–209. https://doi.org/10.1016/j.cognition.2014.10.005

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2012). The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience, 7(4)*, 393–402. https://doi.org/10.1093/scan/nsr005

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58(9)*, 697–720. https://doi.org/10.1037/0003-066X.58.9.697

Kahneman, D., & Frederick, S. (2002). Representativeness Revisited: Attribute Substitution in Intuitive Judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), Heuristics and Biases (1st ed., pp. 49–81). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

Kalla, J. L., & Broockman, D. E. (2020). Reducing Exclusionary Attitudes through Interpersonal Conversation: Evidence from Three Field Experiments. *American Political Science Review, 114(2)*, 410–425. https://doi.org/10.1017/S0003055419000923

Kallioinen, N., Pershina, M., Zeiser, J., Nosrat Nezami, F., Pipa, G., Stephan, A., & König, P. (2019). Moral Judgements on the Actions of Self-Driving Cars and Human Drivers in Dilemma Situations From Different Perspectives. *Frontiers in Psychology, 10*, 2415. https://doi.org/10.3389/fpsyg.2019.02415

Kant, I. (1785). *Grounding for the Metaphysics of Morals*. Hackett

Kant, I. (1797). *The Metaphysical Principles of Virtue*. Hackett

Kleiman, T., Hassin, R. R., & Trope, Y. (2014). The control-freak mind: Stereotypical biases are eliminated following conflict-activated cognitive control. *Journal of Experimental Psychology: General, 143(2)*, 498–503. https://doi.org/10.1037/a0033047

Kleiman, T., Trope, Y., & Amodio, D. M. (2016). Cognitive control modulates attention to food cues: Support for the control readiness model of self-control. *Brain and Cognition, 110*, 94–101. https://doi.org/10.1016/j.bandc.2016.04.006

Koenigs, M., & Tranel, D. (2008). Prefrontal cortex damage abolishes brand-cued changes in cola preference. *Social Cognitive and Affective Neuroscience, 3(1)*, 1–6. https://doi.org/10.1093/scan/nsm032

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature, 446(7138)*, 908–911. https://doi.org/10.1038/nature05631

Kohlberg, L. (1973). The Claim to Moral Adequacy of a Highest Stage of Moral Judgment. *The Journal of Philosophy, 70(18)*, 630. https://doi.org/10.2307/2025030

Kohlberg, L., & Hersh, R. H. (1977). Moral development: A review of the theory. *Theory Into Practice, 16(2)*, 53–59. https://doi.org/10.1080/00405847709542675

Kristol, W. (1994, January 11). How to oppose the health plan-and why. The Wall Street Journal. https://ashbrook.org/viewpoint/onprin-v2n1-kristol/

Laham, S. M., Alter, A. L., & Goodwin, G. P. (2009). Easy on the mind, easy on the wrongdoer: Discrepantly fluent violations are deemed less morally wrong. *Cognition, 112(3)*, 462–466. https://doi.org/10.1016/j.cognition.2009.06.001

Lane, D., & Sulikowski, D. (2017). Bleeding-heart conservatives and hard-headed liberals: The dual processes of moral judgements. *Personality and Individual Differences, 115*, 30–34. https://doi.org/10.1016/j.paid.2016.09.045

Lavie, N., & Dalton, P. (2014). Load Theory of Attention and Cognitive Control (A. C. (Kia) Nobre & S. Kastner, Eds.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199675111.013.003

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load Theory of Selective Attention and Cognitive Control. *Journal of Experimental Psychology: General, 133(3)*, 339–354. https://doi.org/10.1037/0096-3445.133.3.339

Lee, J. J., & Gino, F. (2015). Poker-faced morality: Concealing emotions leads to utilitarian decision making. *Organizational Behavior and Human Decision Processes, 126*, 49–64. https://doi.org/10.1016/j.obhdp.2014.10.006

Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology, 68(1)*, 151–158. https://doi.org/10.1037/0022-3514.68.1.151

Li, Z., Xia, S., Wu, X., & Chen, Z. (2018). Analytical thinking style leads to more utilitarian moral judgments: An exploration with a process-dissociation approach. *Personality and Individual Differences, 131*, 180–184. https://doi.org/10.1016/j.paid.2018.04.046

Lick, D. J., & Johnson, K. L. (2015). The Interpersonal Consequences of Processing Ease: Fluency as a Metacognitive Foundation for Prejudice. *Current Directions in Psychological Science, 24(2)*, 143–148. https://doi.org/10.1177/0963721414558116

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology, 79*, 328–348. https://doi.org/10.1016/j.jesp.2018.08.009

Manfrinati, A., Lotto, L., Sarlo, M., Palomba, D., & Rumiati, R. (2013a). Moral dilemmas and moral principles: When emotion and cognition unite. *Cognition and Emotion, 27(7)*, 1276–1291. https://doi.org/10.1080/02699931.2013.785388

Manfrinati, A., Lotto, L., Sarlo, M., Palomba, D., & Rumiati, R. (2013b). Moral dilemmas and moral principles: When emotion and cognition unite. *Cognition & Emotion, 27(7)*, 1276–1291. https://doi.org/10.1080/02699931.2013.785388

Mansouri, F. A., Tanaka, K., & Buckley, M. J. (2009). Conflict-induced behavioural adjustment: A clue to the executive functions of the prefrontal cortex. *Nature Reviews Neuroscience, 10(2)*, 141–152. https://doi.org/10.1038/nrn2538

McGuire, J., Langdon, R., Coltheart, M., & Mackenzie, C. (2009). A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology, 45(3)*, 577–580. https://doi.org/10.1016/j.jesp.2009.01.002

McKeegan, M. (1993). The politics of abortion: A historical perspective. *Women's Health Issues, 3(3)*, 127–131. https://doi.org/10.1016/S1049-3867(05)80245-2

Mill, J. S. (2014). Utilitarianism. Cambridge University Press. https://doi.org/10.1017/CBO9781139923927

Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience, 24(1)*, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Miller, L. C., Murphy, R., & Buss, A. H. (1981). Consciousness of body: Private and public. *Journal of Personality and Social Psychology, 41(2)*, 397–406. https://doi.org/10.1037/0022-3514.41.2.397

Monin, B., Pizarro, D. A., & Beer, J. S. (2007). Deciding versus Reacting: Conceptions of Moral Judgment and the Reason-Affect Debate. *Review of General Psychology, 11(2)*, 99–111. https://doi.org/10.1037/1089-2680.11.2.99

Montoya, E. R., Terburg, D., Bos, P. A., Will, G.-J., Buskens, V., Raub, W., & van Honk, J. (2013). Testosterone administration modulates moral judgments depending on second-to-fourth digit ratio. *Psychoneuroendocrinology, 38(8)*, 1362–1369. https://doi.org/10.1016/j.psyneuen.2012.12.001

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science, 19(6)*, 549–557. https://doi.org/10.1111/j.1467-9280.2008.02122.x

Moretto, G., Làdavas, E., Mattioli, F., & di Pellegrino, G. (2010). A Psychophysiological Investigation of Moral Judgment after Ventromedial Prefrontal Damage. *Journal of Cognitive Neuroscience, 22(8)*, 1888–1899. https://doi.org/10.1162/jocn.2009.21367

Morsanyi, K., Busdraghi, C., & Primi, C. (2014). Mathematical anxiety is linked to reduced cognitive reflection: A potential road from discomfort in the mathematics classroom to

susceptibility to biases. *Behavioral and Brain Functions, 10(1),* 31. https://doi.org/10.1186/1744-9081-10-31

Ongur, D. (2000). The Organization of Networks within the Orbital and Medial Prefrontal Cortex of Rats, Monkeys and Humans. *Cerebral Cortex, 10(3),* 206–219. https://doi.org/10.1093/cercor/10.3.206

Pastötter, B., Gleixner, S., Neuhauser, T., & Bäuml, K.-H. T. (2013). To push or not to push? Affective influences on moral judgment depend on decision frame. *Cognition, 126(3),* 373–377. https://doi.org/10.1016/j.cognition.2012.11.003

Patil, I. (2015). Trait psychopathy and utilitarian moral judgement: The mediating role of action aversion. *Journal of Cognitive Psychology, 27(3),* 349–366. https://doi.org/10.1080/20445911.2015.1004334

Patil, I., Zucchelli, M. M., Kool, W., Campbell, S., Fornasier, F., Calò, M., Silani, G., Cikara, M., & Cushman, F. (2021). Reasoning supports utilitarian resolutions to moral dilemmas across diverse measures. *Journal of Personality and Social Psychology, 120(2),* 443–460. https://doi.org/10.1037/pspp0000281

Paulhus, D.L., & Williams, K.M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and Psychopathy. *Journal of Research in Personality, 36,* 556 – 563.

Paxton, J. M., Bruni, T., & Greene, J. D. (2014). Are 'counter-intuitive' deontological judgments really counter-intuitive? An empirical reply to Kahane et al. (2012). *Social Cognitive and Affective Neuroscience, 9(9),* 1368–1371. https://doi.org/10.1093/scan/nst102

Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science, 36(1),* 163–177. https://doi.org/10.1111/j.1551-6709.2011.01210.x

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). Everyday Consequences of Analytic Thinking. *Current Directions in Psychological Science, 24(6),* 425–432. https://doi.org/10.1177/0963721415604610

Pennycook, G., & Rand, D. G. (2019). Cognitive Reflection and the 2016 U.S. Presidential Election. *Personality and Social Psychology Bulletin, 45(2),* 224–239. https://doi.org/10.1177/0146167218783192

Piazza, J., & Sousa, P. (2014). Religiosity, Political Orientation, and Consequentialist Moral Thinking. *Social Psychological and Personality Science, 5(3),* 334–342. https://doi.org/10.1177/1948550613492826

Piazza, J., Sousa, P., & Holbrook, C. (2013). Authority dependence and judgments of utilitarian harm. *Cognition, 128(3),* 261–270. https://doi.org/10.1016/j.cognition.2013.05.001

Powell, D., Cheng, P., & Waldmann, M. R. (2016). How should autonomous vehicles behave in moral dilemmas? Human judgments reflect abstract moral principles. Proceedings of the 38th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.

Primi, C., Morsanyi, K., Chiesi, F., Donati, M.A., & Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making, 29(5),* 453-469. Doi: 10.1002/bdm.1883

Rauch, W. A., & Schmitt, K. (2009). Fatigue of Cognitive Control in the Stroop-Task. Proceedings of the annual meeting of the cognitive science society, 31, 750-755.

Reber, R., & Schwarz, N. (1999). Effects of Perceptual Fluency on Judgments of Truth. *Consciousness and Cognition, 8(3),* 338–342. https://doi.org/10.1006/ccog.1999.0386

Reynolds, C. J., Makhanova, A., Ng, B. K. L., & Conway, P. (2020). Bound together for God and country: The binding moral foundations link unreflectiveness with religiosity and political conservatism. *Personality and Individual Differences, 155*, 109632. https://doi.org/10.1016/j.paid.2019.109632

Rolls, E.T. (2005). *Emotion Explained*. Oxford University Press. DOI:10.1093/acprof:oso/9780198570035.001.0001

Royzman, E. B., Landy, J. F., & Leeman, R. F. (2015). Are Thoughtful People More Utilitarian? CRT as a Unique Predictor of Moral Minimalism in the Dilemmatic Context. *Cognitive Science, 39(2),* 325–352. https://doi.org/10.1111/cogs.12136

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as Embodied Moral Judgment. *Personality and Social Psychology Bulletin, 34(8),* 1096–1109. https://doi.org/10.1177/0146167208317771

Schwarz, N., Bless, H., & Bohner, G. (1991). Mood and persuasion: Affective states influence the processing of persuasive communications. *Advances in Experimental Social Psychology, 24,* 161-197. Doi:10.1036/S0065-2601(08)60329-9

Seidel, A., & Prinz, J. (2013). Mad and glad: Musically induced emotions have divergent impact on morals. *Motivation and Emotion, 37(3),* 629–637. https://doi.org/10.1007/s11031-012-9320-7

Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, 114(2), 727–741. https://doi.org/10.1093/brain/114.2.727

Shenhav, A., & Greene, J. D. (2014). Integrative Moral Judgment: Dissociating the Roles of the Amygdala and Ventromedial Prefrontal Cortex. *Journal of Neuroscience, 34(13),* 4741–4749. https://doi.org/10.1523/JNEUROSCI.3390-13.2014

Sidgwick, H. (1962). The Methods of Ethics. Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-81786-3

Siegel, J. Z., & Crockett, M. J. (2013). How serotonin shapes moral judgment and behavior. Annals of the New York Academy of Sciences, 1299(1), 42–51. https://doi.org/10.1111/nyas.12229

Sirota, M., Dewberry, C., Juanchich, M., Valuš, L., & Marshall, A. C. (2020). Measuring cognitive reflection without maths: Development and validation of the verbal cognitive reflection test. *Journal of Behavioral Decision Making, 1-22*. https://doi.org/10.1002/bdm.2213

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119(1)*, 3–22. https://doi.org/10.1037/0033-2909.119.1.3

Song, H., & Schwarz, N. (2008). Fluency and the Detection of Misleading Questions: Low Processing Fluency Attenuates the Moses Illusion. *Social Cognition, 26(6)*, 791–799. https://doi.org/10.1521/soco.2008.26.6.791

Spears, D. F., Okan, Y., Cándido, A., & González, F. (May, 2014). Cognitive Reflection and Callous Affect: Two independent factors related to utilitarian decisions on moral dilemmas. 7th Simposio AME, Granada, 128-134.

Spears, D., Fernández-Linsenbarth, I., Okan, Y., Ruz, M., & González, F. (2018). Disfluent fonts lead to more utilitarian decisions in moral dilemmas. *Psicológica Journal, 39(1)*, 41–63. https://doi.org/10.2478/psicolj-2018-0003

Spears, D., Okan, Y., Hinojosa-Aguayo, I., Perales, J. C., Ruz, M., & González, F. (2021). Can induced reflection affect moral decision-making? *Philosophical Psychology*, 1–19. https://doi.org/10.1080/09515089.2020.1861234

Strohminger, N., Lewis, R. L., & Meyer, D. E. (2011). Divergent effects of different positive emotions on moral judgment. *Cognition, 119(2)*, 295–300. https://doi.org/10.1016/j.cognition.2010.12.012

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18(6)*, 643–662. https://doi.org/10.1037/h0054651

Suter, R.S., & Hertwig, R. (2011). Time and moral judgment. *Cognition, 119,* 454-458. Doi: 10.1016/j.cognition.2011.01.018.

Swami, V., Voracek, M., Stieger, S., Tran, U.S., & Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition, 133,* 572-585. doi: 10.1016/j.cognition.2014.08.006.

Tanne, J. H., & Dyer, O. (2008). Merck used selective data in Vioxx publications, JAMA says. 336, 7.

Thompson, V.A., Prowse Turner, J.A., Pennycook, G., Ball, L.J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and percetual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128,* 237-251. DOI: 10.1016/j.cognition.2012.09.012

Timmons, S., & Byrne, R. M. (2019). Moral fatigue: The effects of cognitive fatigue on moral reasoning. *Quarterly Journal of Experimental Psychology, 72(4),* 943–954. https://doi.org/10.1177/1747021818772045

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning, 20(2),* 147–168. https://doi.org/10.1080/13546783.2013.844729

Tranel, D., & Damasio, H. (1994). Neuroanatomical correlates of electrodermal skin conductance responses. *Psychophysiology, 31(5),* 427–438. https://doi.org/10.1111/j.1469-8986.1994.tb01046.x

Trémolière, B., Neys, W. D., & Bonnefon, J.-F. (2012). Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition, 124(3),* 379–384. https://doi.org/10.1016/j.cognition.2012.05.011

Trémolière, B., Neys, W. D., & Bonnefon, J. F. (2017). Reasoning and moral judgment: A common experimental toolbox. In L. J. Ball & V. A. Thompson (Eds.), Internatoinal handbook of thinking and reasoning (pp. 575-589). Psychology Press.

Trippas, D., Pennycook, G., Verde, M. F., & Handley, S. J. (2015). Better but still biased: Analytic cognitive style and belief bias. *Thinking & Reasoning, 21(4),* 431–445. https://doi.org/10.1080/13546783.2015.1016450

Trippas, D., Thompson, V. A., & Handley, S. J. (2017). When fast logic meets slow belief: Evidence for a parallel-processing model of belief bias. *Memory & Cognition, 45(4),* 539–552. https://doi.org/10.3758/s13421-016-0680-1

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28(2),* 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

Ugazio, G., Lamm, C., & Singer, T. (2012). The role of emotions for moral judgments depends on the type of emotion and moral scenario. *Emotion, 12(3),* 579–590. https://doi.org/10.1037/a0024611

Valdesolo, P., & DeSteno, D. (2006). Manipulations of Emotional Context Shape Moral Judgment. *Psychological Science, 17(6),* 476–477. https://doi.org/10.1111/j.1467-9280.2006.01731.x

Van Berkel, L., Crandall, C. S., Eidelman, S., & Blanchar, J. C. (2015). Hierarchy, Dominance, and Deliberation: Egalitarian Values Require Mental Effort. *Personality and Social Psychology Bulletin, 41(9)*, 1207–1222. https://doi.org/10.1177/0146167215591961

Wheatley, T., & Haidt, J. (2005). Hypnotic Disgust Makes Moral Judgments More Severe. *Psychological Science, 16(10)*, 780–784. https://doi.org/10.1111/j.1467-9280.2005.01614.x

Wiech, K., Kahane, G., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2013). Cold or calculating? Reduced activity in the subgenual cingulate cortex reflects decreased emotional aversion to harming in counterintuitive utilitarian judgment. *Cognition, 126(3)*, 364–372. https://doi.org/10.1016/j.cognition.2012.11.002

Wiegmann, A., Okan, Y., & Nagel, J. (2012). Order effects in moral judgment. *Philosophical Psychology, 25(6)*, 813–836. https://doi.org/10.1080/09515089.2011.631995

Wright, J. C., & Baril, G. (2011). The role of cognitive resources in determining our moral intuitions: Are we all liberals at heart? *Journal of Experimental Social Psychology, 47(5)*, 1007–1012. https://doi.org/10.1016/j.jesp.2011.03.014

Yilmaz, O., & Saribay, S. A. (2017a). The relationship between cognitive style and political orientation depends on the measures used. *Judgment and Decision Making, 12(2)*, 140–147.

Yilmaz, O., & Saribay, S. A. (2017b). Activating analytic thinking enhances the value given to individualizing moral foundations. *Cognition, 165*, 88–96. https://doi.org/10.1016/j.cognition.2017.05.009

Youssef, F. F., Dookeeram, K., Basdeo, V., Francis, E., Doman, M., Mamed, D., Maloo, S., Degannes, J., Dobo, L., Ditshotlo, P., & Legall, G. (2012). Stress alters personal moral decision making. *Psychoneuroendocrinology, 37(4)*, 491–498. https://doi.org/10.1016/j.psyneuen.2011.07.017

# Appendices

# Appendix A: Can induced reflection affect moral decision making?

***Appendix A.1: CRT, CRT NEW and BNT questions and feedback CRT Practice question:***

*CRT Practice:*

If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together? ____ [Correct answer = 4 days; intuitive answer = 9].

*Feedback:*

I'm sorry, your response is not correct. Maybe it would be easier if you imagine that the barrel has a certain capacity. For example, how much would Juan and Mary drink individually each day if the barrel has 12 liters?

*Practice Explanation (for both no feedback and no feedback):*

Many people think that the answer is 9, but, if you look at the question, John would finish the barrel in three days earlier by himself. Therefore, the answer has to be less than 6 days. Further, if you imagine that it's a 12 liter barrel, this means that John has to drink 2 liters of water each day to finish it in 6 days, and Mary needs to drink a liter everyday to finish in 12 days. Therefore, together, they drink 3 liters each day and finish in 4 days.

Day 1: 3 liters drunk, John = 2 liters, Mary = 1 liter

Day 2: 6 liters drunk, John = 2 liters, Mary = 1 liter

Day 3: 9 liters drunk, John = 2 liters, Mary = 1 liter

***CRT questions:***

*CRT 1:*

A bat and a ball costs 1.10 € total. The bat costs one euro more than the ball. How much does the ball cost?

Please, enter the answer and mark it down on your piece of paper too. Use whole numbers, no decimals.

____Cents

[Correct answer = .05 or 5; Intuitive answer = .10 or 10]

*Feedback:*

I'm sorry, your response was not correct. Did you check how much the ball and bat cost together according to the answer you have given?

[Repeat question]

Please, modify your answer now. Remember to use whole numbers, without decimals.

*CRT 2:*

If it takes 5 machines 5 minutes to make 5 objects, how long would it take 100 machines to make 100 widgets?

Please, enter the answer and mark it down on your piece of paper too.

____ Minutes [Correct answer = 5 minutes; Intuitive answer = 100 minutes]

*Feedback:*

I'm sorry, your response is not correct. Have you considered that each machine makes one complete object in your calculations and how much time it takes a machine to make 1 object?

[Repeat question]

Please, modify your response now.

*CRT 3:*

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? Please, enter the answer and mark it down on your piece of paper too.

____days [Correct answer = 47; Intuitive answer = 24]

*Feedback:*

I'm sorry your answer is incorrect. Have you considered the proportion of lily pads that there are in the lake on any day in respect to the day before?

[Repeat question]

Please, modify your response now.

*CRT New*

*CRT New 1:*

If three elves can wrap three toys in an hour, how many elves are needed to wrap six toys in 2 hours?

____ elves [Correct answer = 3 elves; Intuitive answer = 6 elves]

*Feedback:*

I'm sorry, your answer is incorrect. Have you considered how much time it takes an Elf to wrap a toy in your calculations?

[Repeat question]

Please, modify your response now.

*CRT New 2:*

Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are there in the class?

___ students [Correct answer = 29; Intuitive answer = 30]

*Feedback:*

I'm sorry. Your answer is incorrect. Have you considered imagining how many students remain below and above Jerry so that Jerry is always the fifteenth place regardless if you start counting from below or from above.

[Repeat question]

Please, modify your response now.

*CRT New 3:*

In an athletics team, tall members are three times more likely to win a medal than short members. This year the team has won 60 medals so far. How many of these have been won by short athletes?

___ medals [Correct answer = 15; Intuitive answer = 20]

*Feedback:*

I'm sorry, your answer is incorrect. Have you checked that if you multiplied your response by three, then added your answer to the result equals sixty?

[Repeat question]

Please, modify your response now.

*BNT Practice:*

Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws how many times would the die show the number 6? _____out of 70 throws.

[Correct answer = 20]

*Feedback:*

I'm sorry, your response is not correct. Have you thought about how many sides there are in total and how many you have to take account of according to the question?

[Repeat question]

Please, modify your response now. Remember to use whole numbers, without decimals.

*Explanation (for practice):*

To resolve the problem we have to first consider that a 6 sided die that one side has double the probability to land more than the others is really more like a die with seven sides. This is because if we roll the die an infinite amount of times, 6 will be rolled two times more than any other side. It is as though the 6 occupies two places at once; as though it would have been an extra number. If this is true, the probability of each number, except for 6, would be 1/7, while the number 6 would be 2/7.

Therefore, if we roll a die 70 times, each number should be rolled 10 times, except 6 which would be 20.

Numbers 1 to 5: 70 * 1/7 = 10 times each (in total, these numbers would be rolled 50 times); Number 6: 70 * 2/7 = 20 times

***BNT questions:***

*BNT 1:*

Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)? _____out of 50 throws.

[Correct answer = 30]

*Feedback:*

I’m sorry your answer is incorrect. Did you consider how many faces the die has in total and how many you have to take into account according to the question?

[Repeat question]

Please modify your answer

*BNT 2:*

Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? Please indicate the probability in percent. _____ %

[Correct answer = 25]

*Feedback:*

I’m sorry your answer is incorrect. Have you considered how many men there are in the village and how many of them are in the chorus?

[Repeat question]

Please, modify your answer.

*BNT 3:*

In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red? _____%

[Correct answer = 50]

*Feedback:*

I’m sorry, your answer is incorrect. Have you considered the proportion of red poisonous mushrooms over the total number of poisonous mushrooms supposing that there are 1000 mushrooms in the forest?

[Repeat question]

Please, modify your answer.

***Appendix A.2: Dilemmas***

*Walnuts (Practice):*

A girl decided to make a chocolate cake. She looks in a recipe book and finds one that calls for a cup of walnuts. She does not like walnuts, but she does like almonds. Unfortunately, she has the right amount of walnuts and almonds; therefore she can make the cake with either nut.

Do you think that the girl should substitute the walnuts for almonds in order to avoid eating walnuts?

Definitely No Definitely Yes

1 2 3 4 5 6

*Crying Baby:*

Enemy soldiers have taken over your village. They have orders to kill all they encounter. You and some of your townspeople have sought refuge in the basement of a large house. Outside you hear the voices of soldiers who have come to search the house for valuables. Your baby, who is with you, begins to cry loudly. You put your hand over his mouth to block the sound of its cries. If you uncover the baby's mouth, its cries will call attention to the soldiers, who will kill everyone they find in the basement, including you and your baby. In order to save yourself and the others you must keep your hand on his mouth and smother your baby to death.

Do you think you should smother your child in order to save yourself and the other hidden townspeople?

Definitely No Definitely Yes

1 2 3 4 5 6

*Submarine:*

You are a crewperson on a marine research submarine traveling underneath a large iceberg. An on board explosion has damaged the ship, and it's losing the majority of its oxygen. One of the sailors is injured is quickly losing blood. The injured sailor will die from his injuries regardless of what happens. The remaining oxygen is not sufficient for the whole team to make it to the surface. The only way to save the rest of the crew is to shoot the injured sailor in order for there to be enough oxygen for the rest of the crew to survive.

Do you think you should shoot the sailor to save yourself and the rest of the crew?

Definitely No Definitely Yes

1 2 3 4 5 6

*Rescue 911:*

You are the sole paramedic riding on a rescue helicopter, responding to a devastating flood. You have rescued several critically injured people and you have been treating them to keep them alive. Suddenly, the pilot notifies you that there has been an engine malfunction and the helicopter can no longer support the weight of you, the pilot, and all the injured people on board. If the helicopter crashes, it will kill everyone on board including you, the pilot, and the injured people that have rescued. In order to avoid a crash that will kill everyone on the helicopter, you realize that you must lighten the load enough to keep the helicopter aloft. Because all f the equipment is bolted down, and the injured people need you, the only way to do this is to throw one of the injured people

off the helicopter. This will lighten the load enough to keep the helicopter in the air but will cause the death of the person thrown off. Do you think you should throw one of the injured people from the helicopter in order to save yourself and everyone else on board?

Definitely No Definitely Yes

1 2 3 4 5 6

***Appendix A.3: Previous knowledge question***

Thank you for your participation! It's important for the objectives of this investigation that your responses only reflect your personal judgments and estimations close to the questions that were presented to you. Thus we ask that if you have used any additional information (class notes, internet searches, consulting other people, etc…) or, for example, you had previously seen the answers to the mathematical tasks, simply mark the option 'yes'. In this case, please indicate in the commentary window, which will appear next page, what you remember in respect to those questions. On the contrary, mark the option 'no'. In either case, this response will not have any repercussions to you, but this information will help give us an idea about the validity about our data. Thank you.

*In Sample B*

Did you have any previous knowledge of the mathematical questions? Yes or No

Did you previously know one (or more) of the stories? Yes or No

# **Appendix B:** Disfluent fonts lead to more utilitarian decisions in moral dilemmas

**Appendix B.1: Dilemmas**

Practice dilemmas (Greene et al., 2004), slightly modified to make them more familiar to Spanish participants.

*Brownies*

A girl decided to make a chocolate cake. She searches in a recipe book finding one that calls for a cup of nuts. However, she does not like nuts, but she does almonds. Luckily, she has got enough nuts and almonds in the cupboard, so she could make the cake with either.

Would it be appropriate for her to substitute nuts for almonds in order to avoid eating nuts?

*Standard Turnips*

A farm worker is driving a wheat-harvesting machine approaching a fork in the road. By choosing the path on the left, he will harvest ten kilograms of wheat, whereas by choosing the path on the right he will harvest twenty kilograms. If he does nothing, the wheat-harvesting machine will turn to the left.

Would it be it appropriate for him to turn the wheat-harvesting machine to the right in order to harvest twenty kilograms of wheat instead of ten?

Fillers (Moore et al., 2008)

*Stock Tip*

You are a managing consultant working on a case for a large corporate client. You have access to confidential information that would be very useful to investors. You also have a friend who plays

the stock market and to whom you owe a sizable sum of money. By providing her with certain confidential information you could help her make a lot of money, considerably more that you owe her. If you do this, she will cancel your debt. However releasing information could cause other investors to lose a great deal of money

Would it be appropriate for you to release this information to your friend so that she will cancel your debt?

*Taxes*

You are the owner of a small business. Your company is making ends meet, but it is not terribly profitable. You would really like to buy a new car. However its price is beyond your price range. It occurs to you that you could lower your taxes by pretending that some of your personal expenses are business expenses. For example, you could say that the stereo in your bedroom is being used in the lounge at the office, or that dinners out with your wife are dinners with clients. This lie would allow you to hide enough money from the IRS to buy the new car.

Would it be appropriate for you to pretend that certain personal expenses are business expenses in order to buy a new car?

Moral Dilemmas (Moore et al., 2008)

*Burning Building – Personal Version*

You and five other people are trapped in a burning building. There is only one emergency exit through which all of you could escape to safety, but it is blocked by burning debris. You notice another person in the hallway leading to the exit. This person has a mortal wound and is blocking

a small hole at the exit door. You and the five people behind you do not have time to climb through the small hole. You realize that you could grab the injured man and use his body as a battering-ram to break through the debris blockage in the hallway that is preventing your scape. Doing this is certain to kill him, although if you do not do it, you and the five people behind you will die.

Would it be appropriate for you to kill this person who is badly injured in order to save yourself and the other five people?

*Burning Building – Impersonal Version*

You and five other people are trapped in a burning building. There is only one emergency exit through which all of you could escape to safety, but it is blocked by burning debris. You notice another person in the hallway leading to the exit. This person has a mortal wound and is blocking a small hole at the exit door. You and the five people behind you do not have time to climb through the small hole. The hallway's emergency system puts out fire by eliminating oxygen from the hall and you can activate the system by pressing a button. By doing that, the fire will go out, but the injured person will suffocate and die. However, if you do not do it, you and the five people behind you will die.

Would it be appropriate for you to activate the system in order to save yourself and the other five people?

*Crying Baby – Personal Version*

Enemy soldiers have taken over your village. They have orders to kill all remaining civilians. You and some of your townspeople have found refuge in the cellar of a large house. Outside you hear the voices of soldiers who have come to search the house for valuables. Your baby, who is with

you in the room, begins to cry loudly. You put your hand over his mouth to block the sound of his crying. If you remove your hand from his mouth his crying will summon the attention of the soldiers who will kill all of you, including you and your baby. To save yourself and the others you must smother your baby to death.

Would it be appropriate for you to smother your baby in order to save yourself and the other townspeople who are hiding?

*Crying Baby – Impersonal Version*

Enemy soldiers have taken over your village. They have orders to kill all remaining civilians. You and some of your townspeople have found refuge in the cellar of a large house. Outside you hear the voices of soldiers who have come to search the house for valuables. Your baby, who is with you in the room, begins to cry loudly. His crying will summon the attention of the soldiers who will kill all of you, including you and your baby. There is a furnace which is very noisy when it is working. If you turn it on, its noise would block the sound of the crying, but the room will become deadly hot for the baby and he will die.

Would it be appropriate for you to turn on the furnace in order to save yourself and the other townspeople who are hiding?

# **Appendix C:** Do cognitive control tasks increase utilitarian decisions on moral dilemmas?

# Appendix C.1: Stroop experiment

## *Appendix C.1.1: Dilemmas*

### *Practice Dilemmas*

*Brownies*

A girl decided to make a chocolate cake. She searches in a recipe book finding one that calls for a cup of nuts. However, she does not like nuts, but she does almonds. Luckily, she has got enough nuts and almonds in the cupboard, so she could make the cake with either. Would it be appropriate for her to substitute nuts for almonds in order to avoid eating nuts?

*Standard Turnips*

A farm worker is driving a wheat-harvesting machine approaching a fork in the road. By choosing the path on the left, he will harvest ten kilograms of wheat, whereas by choosing the path on the right he will harvest twenty kilograms. If he does nothing, the wheat-harvesting machine will turn to the left. Would it be it appropriate for him to turn the wheat-harvesting machine to the right in order to harvest twenty kilograms of wheat instead of ten?

### *Filler dilemmas:*

*Taxes*

You are the owner of a small business. Your company is making ends meet, but it is not terribly profitable. You would like very much to buy the new sports car that you saw your neighbor driving the other day. However, it is beyond your price range. It occurs to you that you could lower your taxes by pretending that some of your personal expenses are business expenses. For example, you

could pretend that the stereo in your bedroom is being used in the lounge at the office, or that dinners out with your wife are dinners with clients. This would allow you to hide enough money from the IRS to buy the new car. Is it appropriate for you to pretend that certain personal expenses are business expenses in order to buy a new car?

*Stock Tip*

You are the owner of a small business. Your company is making ends meet, but it is not terribly profitable. You would like very much to buy the new sports car that you saw your neighbor driving the other day. However, it is beyond your price range. It occurs to you that you could lower your taxes by pretending that some of your personal expenses are business expenses. For example, you could pretend that the stereo in your bedroom is being used in the lounge at the office, or that dinners out with your wife are dinners with clients. This would allow you to hide enough money from the IRS to buy the new car. Is it appropriate for you to pretend that certain personal expenses are business expenses in order to buy a new car?

***Dilemmas:***

*Burning Building – Personal Version*

You and five other people are trapped in a burning building. There is only one emergency exit through which all of you could escape to safety, but it is blocked by burning debris. You notice another person in the hallway leading to the exit. This person has a mortal wound and is blocking a small hole at the exit door. You and the five people behind you do not have time to climb through the small hole. You realize that you could grab the injured man and use his body as a battering-

ram to break through the debris blockage in the hallway that is preventing your scape. Doing this is certain to kill him, although if you do not do it, you and the five people behind you will die.

Would it be appropriate for you to kill this person who is badly injured in order to save yourself and the other five people?

*Burning Building – Impersonal Version*

You and five other people are trapped in a burning building. There is only one emergency exit through which all of you could escape to safety, but it is blocked by burning debris. You notice another person in the hallway leading to the exit. This person has a mortal wound and is blocking a small hole at the exit door. You and the five people behind you do not have time to climb through the small hole. The hallway's emergency system puts out fire by eliminating oxygen from the hall and you can activate the system by pressing a button. By doing that, the fire will go out, but the injured person will suffocate and die. However, if you do not do it, you and the five people behind you will die.

Would it be appropriate for you to activate the system in order to save yourself and the other five people?

*Rescue 911 - Personal:*

You are the sole paramedic riding on a rescue helicopter, responding to a devastating flood. You have rescued several critically injured people and you have been treating them to keep them alive. Suddenly, the pilot notifies you that there has been an engine malfunction and the helicopter can no longer support the weight of you, the pilot, and all the injured people on board. If the helicopter crashes, it will kill everyone on board including you, the pilot, and the injured people that have

rescued. In order to avoid a crash that will kill everyone on the helicopter, you realize that you must lighten the load enough to keep the helicopter aloft. Because all f the equipment is bolted down, and the injured people need you, the only way to do this is to throw one of the injured people off the helicopter. This will lighten the load enough to keep the helicopter in the air but will cause the death of the person thrown off. Do you think you should throw one of the injured people from the helicopter in order to save yourself and everyone else on board?

*Rescue 911 – Impersonal:*

You are the sole paramedic riding on a rescue helicopter, responding to a devastating flood. You have rescued several critically injured people and you have been treating them to keep them alive. Suddenly, the pilot notifies you that there has been an engine malfunction and the helicopter can no longer support the weight of you, the pilot, and all the injured people on board. If the helicopter crashes, it will kill everyone on board including you, the pilot, and the injured people that you have rescued. In order to avoid a crash that will kill everyone on the helicopter, you realize that you must lighten the load enough to keep the helicopter aloft. Because all of the equipment is bolted down, and you are needed by the injured people, the only way to do this is to cut the cable that is raising up the last rescued person. This will lighten the load enough to keep the helicopter in the air, but it will also kill the person who is dropped. Is it appropriate for you to kill this injured person in order to save yourself and everyone else on board?

***Appendix C.1.2: Mood Scale***

Please, indicate your mood below with the following scale, selecting one of the options between 1 and 7, where 1 = very sad; 4 = neither sad, nor happy; 7 = very happy. You can respond using the keyboard or the mouse.

1 2 3 4 5 6 7

***Appendix C.1.3: Content questions***

Question (Q): Where was the equipment attached to in the helicopter?

Answer (A):The floor

Q: How was the emergency system activated to put out the fire in the story of the burning building?

A: A button

Q: What was the new thing the person wanted to buy in the story about tax fraud?

A: A car

Q: With whom did the person owe a debt to in the stock market story?

A: a friend

Q: What was needed to be cut to save you and five other people in the helicopter story?

A: A cable

Q: What was needed to be used as a battering ram in order to exit the burning building?

A: A body

# Appendix C.2: Flanker Experiment

### *Appendix C.2.1: Dilemmas*

*Dilemmas A:*

*Rescue 911:*

You are the sole paramedic riding on a rescue helicopter, responding to a devastating flood. You have rescued several critically injured people and you have been treating them to keep them alive. Suddenly, the pilot notifies you that there has been an engine malfunction and the helicopter can no longer support the weight of you, the pilot, and all the injured people on board. If the helicopter crashes, it will kill everyone on board including you, the pilot, and the injured people that have rescued. In order to avoid a crash that will kill everyone on the helicopter, you realize that you must lighten the load enough to keep the helicopter aloft. Because all f the equipment is bolted down, and the injured people need you, the only way to do this is to throw one of the injured people off the helicopter. This will lighten the load enough to keep the helicopter in the air but will cause the death of the person thrown off. Do you think you should throw one of the injured people from the helicopter in order to save yourself and everyone else on board?

*Bus Plunge:*

You are the bus driver for a kindergarten field trip to the zoo. On the way, faulty tires cause the bus to overturn and plunge off of a bridge and into a roaring river. You and three of the children are still in the sinking bus, but the rest of the passengers have been swept away down the river to their deaths. You grab the two children nearest to you and begin to swim toward the exit door. The third remaining child grabs onto your leg. You realize that you are not strong enough to fight the current and swim with all three children holding on to you. The only way to reach the surface

before you and the children drown is to shake the third child off of your leg. This will allow you to bring the two children to the surface with you, but the third child will drown. Is it appropriate for you to shake the child off your leg in order to save yourself and the two other children?

*Space Station:*

You are an engineer on the international space station, in orbit around the Earth. Suddenly, a fire breaks out in the cargo bay. The automatic fire safety system will open the outer door of the cargo bay, letting the oxygen out of the bay and thus putting out the fire. The automatic system only works when the inner portal to the rest of the station has been sealed. One member of the crew, a mechanic in the cargo bay, has put on half of his bulky space suit when the fire begins. He runs to the portal to climb through to safety, but the portal is too narrow to fit through while wearing a space suit. You realize that if he tries to climb through, he will get stuck and the fire will spread through the rest of the station because the safety system will not operate with the portal open. There is no time for him to take off the suit before the fire spreads. The only way you can prevent the spread of the fire, and save your life and the rest of the crew, is to push the mechanic back through the portal and into the cargo bay. This way, the safety system will seal the door and decompress the cargo bay, but it will also suck the mechanic into space and kill him. Is it appropriate for you to push the mechanic back through the portal, and into space, in order to save yourself and the rest of the crew?

*Space station:*

You are the inspector of a nuclear power plant that you suspect has not met its safety requirements.

The plant foreman and you are touring the facility when one of the nuclear fuel rods overheats. The emergency coolant system fails to activate, and a chain reaction is about to begin which will result in a nuclear meltdown. This will release lethal radiation into the nearby town, killing thousands of people. You realize that the only way to stop the meltdown is to push the foreman into the fuel rod assembly. This will remove just enough heat energy from the rod assembly to prevent the nuclear chain reaction. However, it will also incinerate the foreman instantly. Is it appropriate for you to push and kill the foreman in order to save yourself and the nearby town?

*Dilemmas B:*

*Burning Building:*

You and five other people are trapped in a burning building. There is only one emergency exit through which all of you could escape to safety, but it is blocked by burning debris. You notice another person in the hallway leading to the exit. This person has a mortal wound and is blocking a small hole at the exit door. You and the five people behind you do not have time to climb through the small hole. You realize that you could grab the injured man and use his body as a battering-ram to break through the debris blockage in the hallway that is preventing your scape. Doing this is certain to kill him, although if you do not do it, you and the five people behind you will die. Would it be appropriate for you to kill this person who is badly injured in order to save yourself and the other five people?

*Cliffhanger:*

You are a construction worker. You and your crew are high on a scaffold, working on a skyscraper. Suddenly, the scaffolding partially collapses. You and several others are hanging on to a dangling

crossbar, but it cannot hold all of your weight. The worker next to you slips off the crossbar and grabs a hold of your one free arm. However, you realize that the entire structure is about to give way. The only way to avoid you and everyone else falling to your deaths is to repeatedly kick your coworker in the stomach and chest until he lets go of your hand and falls to his certain death. This will remove just enough weight that the rest of you can make it to safety before the scaffold collapses. Is it appropriate for you to kill your co-worker in order to save yourself and the rest of the crew?

*Nuclear Reactor:*

You are the inspector of a nuclear power plant that you suspect has not met its safety requirements. The plant foreman and you are touring the facility when one of the nuclear fuel rods overheats. The emergency coolant system fails to activate, and a chain reaction is about to begin which will result in a nuclear meltdown. This will release lethal radiation into the nearby town, killing thousands of people. You realize that the only way to stop the meltdown is to push the foreman into the fuel rod assembly. This will remove just enough heat energy from the rod assembly to prevent the nuclear chain reaction. However, it will also incinerate the foreman instantly. Is it appropriate for you to push and kill the foreman in order to save yourself and the nearby town?

***Appendix C.2.2: Mood Scale***

Please, indicate your mood below with the following scale, selecting one of the options between 1 and 7, where 1 = very sad; 4 = neither sad, nor happy; 7 = very happy. You can respond using the keyboard or the mouse.

1 2 3 4 5 6 7

***Appendix C.2.3: Content questions***

Question (Q): Where was the equipment attached to in the helicopter?

Answer (A):The floor

Q: What was needed to be used as a battering ram in order to exit the burning building?

A: A body

Q: What caused the bus crash?

A: Tire (blow out)

Q: Where were you working when the scaffolding began to break?

A: Sky Scraper

Q: How do you extinguish the fire on the space station?

A:Removing oxygen

Q: Who do you have to push onto the nuclear rods in order to avoid a nuclear meltdown?

A: The director

**Appendix D:** Cognitive reflection and callous affect: Tow independent factors related to utilitarian decisions on moral dilemmas

# *Appendix D.1: Dilemmas*

*Non-moral dilemmas:*

*Brownies*

A girl decided to make a chocolate cake. She searches in a recipe book finding one that calls for a cup of nuts. However, she does not like nuts, but she does almonds. Luckily, she has got enough nuts and almonds in the cupboard, so she could make the cake with either. Would it be appropriate for her to substitute nuts for almonds in order to avoid eating nuts?

*Standard Turnips*

A farm worker is driving a wheat-harvesting machine approaching a fork in the road. By choosing the path on the left, he will harvest ten kilograms of wheat, whereas by choosing the path on the right he will harvest twenty kilograms. If he does nothing, the wheat-harvesting machine will turn to the left. Would it be it appropriate for him to turn the wheat-harvesting machine to the right in order to harvest twenty kilograms of wheat instead of ten?

*Impersonal dilemmas:*

*Standard Trolley:*

You are at the wheel of a runaway trolley quickly approaching a fork in the tracks. On the tracks extending to the left is a group of five railway workmen. On the tracks extending to the right is a single railway workman. If you do nothing, the trolley will proceed to the left, causing the deaths of these workmen is to hit a switch on your dashboard that will cause the trolley to proceed to the right, causing the death of the single workman. Is it appropriate for you to hit the switch in order to avoid the deaths of the five workmen?

*Standard fumes:*

You are the late-night watchman in a hospital. Due to an accident in the building next door, there are deadly fumes rising up through the hospital's ventilation system. In a certain room of the hospital are three patients. In another room there is a single patient. If you do nothing the fumes will rise up into the room containing the three patients and cause their deaths. The only way to avoid the deaths of these patients is to hit a certain switch, which will cause the fumes to bypass the room containing the three patients. As a result of doing this the fumes will enter the room containing the single patient cause his death. Is it appropriate for you to hit the switch in order to avoid the deaths of the five patients?

*Personal, Low-Conflict:*

*Transplant:*

You are a doctor. You have five patients, each of whom is about to die due to a failing organ of some kind. You have another patient who is healthy. The only way that you can save the lives of the first five patients is to transplant five of this young man's organs (against his will) into the bodies of the other five patients. If you do this, the young man will die, but the other five patients will live. Is it appropriate for you to perform this transplant in order to save five of your patients?

*Plane crash:*

Your plane has crashed in the Himilaywas. The only survivors are yourself, another man, and a young boy. The three of you travel for days, battling extreme cold and wind. Your only chance at survival is to find your way to a small village on the other side of the mountain, several days away. The boy has a broken leg and cannot move very quickly. His chances of surviving the journey are

essentially zero. Without food, you and the other man will probably die as well. The other man suggests that you sacrifice the boy and eat his remains over the next few days. Would you kill this boy so that you and the other man may survive your journey?

*Personal, High-conflict:*

*Footbridge:*

A runaway trolley is heading down the tracks toward five workmen who will be killed if the trolley proceeds on its present course. You are on a footbridge over the tracks, in between the approaching trolley and the five workmen. Next to you on this footbridge is a stranger who happens to be very large. The only way to save the lives of the five workmen is to push this stranger off the bridge and onto the tracks below where his large body will stop the trolley. The stranger will die if you do this, but the workman will be saved. Is it appropriate for you to push the stranger on to the tracks in order to save the five workmen?

*Crying baby:*

Enemy soldiers have taken over your village. They have orders to kill all remaining civilians. You and some of your townspeople have sought refuge in the cellar of a large house. Outside you hear the voices of soldiers who have come to search the house for valuables. Your baby begins to cry loudly. You cover his mouth to block the sound. If you remove your hand from his mouth his crying will summon the attention of the soldiers who will kill you, your child, and the others hiding out in the cellar. To save yourself and the others you must smother your child to death. Is it appropriate for you to smother your child in order to save yourself and the other townspeople?

***Appendix D.2:*** **Self-Reported Psychopathy Scale III (SRP-III)**

Please rate the degree to which you agree with the following statements about you. You can be honest because your name will be detached from the answers as soon as they are submitted.

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Disagree Strongly | Disagree | Neutral | Agree | Agree Strongly |

1. I'm a rebellious person.
2. I'm more tough-minded than other people.
3. Most people are wimps.
4. I've often done something dangerous just for the thrill of it.
5. It tortures me to see an animal injured.
6. I have pretended to be someone else in order to get something.
7. I like to see fist-fights.
8. My friends would say that I am a warm person.
9. I would get a kick out of 'scamming' someone.
10. I avoid horror movies.
11. I feel so sorry when I see a homeless person.
12. It's fun to see how far you can push people before they get upset.
13. I enjoy doing wild things.

14. I don’t bother to keep in touch with my family anymore.

15. I rarely follow the rules.

16. I never cry at movies.

17. You should take advantage of other people before they do it to you.

18. People sometimes say that I’m cold-hearted.

19. I love violent sports and movies.

20. Sometimes you have to pretend you like people to get something out of them.

21. I’m a soft-hearted person.

22. People are too sensitive when I tell them the truth about themselves.

23. I keep getting in trouble for the same things over and over.

24. People cry way too much at funerals.

25. You can get what you want by telling people what they want to hear.

26. I easily get bored.

27. I never feel guilty over hurting others.

28. A lot of people are “suckers” and can easily be fooled.

29. I admit that I often “mouth off” without thinking.

30. I sometimes dum friends that I don’t need anymore.

***Appendix D.2:*** **CRT**

*CRT 1:*

A bat and a ball costs 1.10 € total. The bat costs one euro more than the ball. How much does the ball cost?

Please, enter the answer and mark it down on your piece of paper too. Use whole numbers, no decimals.

____Cents

[Correct answer = .05 or 5; Intuitive answer = .10 or 10]

*CRT 2:*

If it takes 5 machines 5 minutes to make 5 objects, how long would it take 100 machines to make 100 widgets?

Please, enter the answer and mark it down on your piece of paper too.

____ Minutes [Correct answer = 5 minutes; Intuitive answer = 100 minutes]

*CRT 3:*

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Please, enter the answer and mark it down on your piece of paper too.

____days [Correct answer = 47; Intuitive answer = 24]

# **Appendix E:** Political and Economic beliefs, deliberative thinking, and moral decisions

# *Appendix E.1: Dilemmas*

*Rescue 911:*

You are the sole paramedic riding on a rescue helicopter, responding to a devastating flood. You have rescued several critically injured people and you have been treating them to keep them alive. Suddenly, the pilot notifies you that there has been an engine malfunction and the helicopter can no longer support the weight of you, the pilot, and all the injured people on board. If the helicopter crashes, it will kill everyone on board including you, the pilot, and the injured people that have rescued. In order to avoid a crash that will kill everyone on the helicopter, you realize that you must lighten the load enough to keep the helicopter aloft. Because all f the equipment is bolted down, and the injured people need you, the only way to do this is to throw one of the injured people off the helicopter. This will lighten the load enough to keep the helicopter in the air but will cause the death of the person thrown off. Do you think you should throw one of the injured people from the helicopter in order to save yourself and everyone else on board?

*Bus Plunge:*

You are the bus driver for a kindergarten field trip to the zoo. On the way, faulty tires cause the bus to overturn and plunge off of a bridge and into a roaring river. You and three of the children are still in the sinking bus, but the rest of the passengers have been swept away down the river to their deaths. You grab the two children nearest to you and begin to swim toward the exit door. The third remaining child grabs onto your leg. You realize that you are not strong enough to fight the current and swim with all three children holding on to you. The only way to reach the surface before you and the children drown is to shake the third child off of your leg. This will allow you

to bring the two children to the surface with you, but the third child will drown. Is it appropriate for you to shake the child off your leg in order to save yourself and the two other children?

*Space Station:*

You are an engineer on the international space station, in orbit around the Earth. Suddenly, a fire breaks out in the cargo bay. The automatic fire safety system will open the outer door of the cargo bay, letting the oxygen out of the bay and thus putting out the fire. The automatic system only works when the inner portal to the rest of the station has been sealed. One member of the crew, a mechanic in the cargo bay, has put on half of his bulky space suit when the fire begins. He runs to the portal to climb through to safety, but the portal is too narrow to fit through while wearing a space suit. You realize that if he tries to climb through, he will get stuck and the fire will spread through the rest of the station because the safety system will not operate with the portal open. There is no time for him to take off the suit before the fire spreads. The only way you can prevent the spread of the fire, and save your life and the rest of the crew, is to push the mechanic back through the portal and into the cargo bay. This way, the safety system will seal the door and decompress the cargo bay, but it will also suck the mechanic into space and kill him. Is it appropriate for you to push the mechanic back through the portal, and into space, in order to save yourself and the rest of the crew?

*Crying Baby*

Enemy soldiers have taken over your village. They have orders to kill all they encounter. You and some of your townspeople have sought refuge in the basement of a large house. Outside you

hear the voices of soldiers who have come to search the house for valuables. Your baby, who is with you, begins to cry loudly. You put your hand over his mouth to block the sound of its cries. If you uncover the baby's mouth, its cries will call attention to the soldiers, who will kill everyone they find in the basement, including you and your baby. In order to save yourself and the others you must keep your hand on his mouth and smother your baby to death. Do you think you should smother your child in order to save yourself and the other hidden townspeople?

*Submarine:*

You are a crewperson on a marine research submarine traveling underneath a large iceberg. An on board explosion has damaged the ship, and it's losing the majority of its oxygen. One of the sailors is injured is quickly losing blood. The injured sailor will die from his injuries regardless of what happens. The remaining oxygen is not sufficient for the whole team to make it to the surface. The only way to save the rest of the crew is to shoot the injured sailor in order for there to be enough oxygen for the rest of the crew to survive. Do you think you should shoot the sailor to save yourself and the rest of the crew?

## *Appendix E.2: Moral Foundations Questionnaire (MFQ)*

*Part 1. When you decide whether something is right or wrong, to what extent are the following considerations relevant to your thinking? Please rate each statement using this scale:*

[0] = not at all relevant (This consideration has nothing to do with my judgments of right and wrong)

[1] = not very relevant

[2] = slightly relevant

[3] = somewhat relevant

[4] = very relevant

[5] = extremely relevant (This is one of the most important factors when I judge right and wrong)

1. Whether or not someone suffered emotionally
2. Whether or not some people were treated differently than others
3. Whether or not someone's action showed love for his or her country
4. Whether or not someone showed a lack of respect for authority
5. Whether or not someone violated standards of purity and decency
6. Whether or not someone was good at math
7. Whether or not someone cared for someone weak or vulnerable
8. Whether or not someone acted unfairly
9. Whether or not someone did something to betray his or her group
10. Whether or not someone conformed to the traditions of society

11. Whether or not someone did something disgusting

12. Whether or not someone was cruel

13. Whether or not someone was denied his or her rights

14. Whether or not someone showed a lack of loyalty

15. Whether or not an action caused chaos or disorder

16. Whether or not someone acted in a way that God would approve of

*Part 2. Please read the following sentences and indicate your agreement or disagreement:*

| [0] | [1] | [2] | [3] | [4] | [5] |
|---|---|---|---|---|---|
| Strongly disagree | Moderately disagree | Slightly disagree | Slightly agree | Moderately agree | Strongly agree |

17. Compassion for those who are suffering is the most crucial virtue.

18. When the government makes laws, the number one principle should be ensuring that everyone is treated fairly.

19. I am proud of my country's history.

20. Respect for authority is something all children need to learn.

21. People should not do things that are disgusting, even if no one is harmed.

22. It is better to do good than to do bad.

23. One of the worst things a person could do is hurt a defenseless animal.

24. Justice is the most important requirement for a society.

25. People should be loyal to their family members, even when they have done something wrong.

26. Men and women each have different roles to play in society.

27. I would call some acts wrong on the grounds that they are unnatural.

28. It can never be right to kill a human being.

29. I think it's morally wrong that rich children inherit a lot of money while poor children inherit nothing.

30. It is more important to be a team player than to express oneself.

31. If I were a soldier and disagreed with my commanding officer's orders, I would obey anyway because that is my duty.

32. Chastity is an important and valuable virtue.

# *Appendix E.3: Political-economic measures*

*Appendix E.3.1.: Self-reported Political scale*

When it comes to politics, do you consider yourself to be progressive, moderate, conservative or another option.

*Appendix E.3.2.: Social and Economic Conservatism Scale (SECS)*

Please indicate the degree to how positive or negative you feel about each issue on the scale of 0 to 100, where 0 indicates the highest degree of unfavorable position and 100 indicates the highest degree of favorable position, and 50 indicates a neutral position on the issue.

1. Abortion
2. Limited government
3. Military and national security
4. Religion
5. Welfare benefits
6. Gun ownership
7. Traditional marriage
8. Traditional values
9. Fiscal responsibility
10. Business
11. The family unit
12. Patriotism

*Appendix E.3.3: Support for the Free Market*

Please indicate to degree which demonstrates your agreement or disagreement of the following topics using a punction from 1 to 4, where 1 indicates the highest degree of disagreement, and 4 the highest degree of agreement.

1. An economic system based on free markets unrestrained by government interference automatically works best to meet human needs.
2. I support the free-market system, but not at the expense of environmental quality.
3. The free-market system may be efficient for resource allocation, but it is limited in its capacity to promote social justice.
4. The preservation of the free market system is more important than localized environmental concerns.
5. Free and unregulated markets pose important threats to sustainable development.
6. The free-market system is likely to promote unsustainable consumption.

# *Appendix E.4: Deliberative Measures*

*Appendix E.4.1: CRT New*

*CRT New 1:*

If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together?

____ [Correct answer = 4 days; intuitive answer = 9].

*CRT New 2:*

If three elves can wrap three toys in an hour, how many elves are needed to wrap six toys in 2 hours?

____ elves [Correct answer = 3 elves; Intuitive answer = 6 elves]

*CRT New 3:*

Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are there in the class?

___ students [Correct answer = 29; Intuitive answer = 30]

*Appendix E.4.2: BNT*

*BNT 1:*

Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)? _____out of 50 throws.

[Correct answer = 30]

*BNT 2:*

Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? Please indicate the probability in percent. _____ %

[Correct answer = 25]

*BNT 3:*

Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws how many times would the die show the number 6? _____out of 70 throws.

[Correct answer = 20]

*BNT 4:*

In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red? _____%

*Appendix E.4.3: Belief Bias Syllogisms*

*Conflict syllogisms*

Everything that has a motor needs oil

Cars need oil

Therefore, cars have a motor (Invalid – believable)

Does the conclusion follow from the premises?

All unemployed people are poor

Marta is not unemployed

Therefore, Marta is not poor (Invalid – believable)

Does the conclusion follow from the premises?

All mammals can walk

Whales are mammals

Therefore, whales can walk (valid – unbelievable)

Does the conclusion follow from the premises?

All animals like water

Cats do not like water

Therefore, cats are not animals (valid – unbelievable)

Does the conclusion follow from the premises?

*No conflict syllogisms:*

All pistols are dangerous

Swords are dangerous

Therefore, swords are pistols (invalid – unbelievable)

Does the conclusion follow from the premises?

Everything made from wood can be used for fire

Gasoline is not made from wood

Therefore, gasoline cannot be used for fire (invalid – unbelievable)

Does the conclusion follow from the premises?

All birds have feathers

Eagles are birds

Therefore, eagles have feathers (valid – believable)

Does the conclusion follow from the premises?

All cows have four hooves

Snakes do not have any hooves

Therefore, snakes are not cows (valid – believable)

Does the conclusion follow from the premises?