TESIS DOCTORAL

THE ROLE OF RACE AND EMOTIONAL EXPRESSIONS ON TRUST DECISIONS

DOCTORANDA

MARÍA TORTOSA MOLINA

DIRECTORES

MARÍA RUZ CÁMARA Y JUAN LUPIÁÑEZ CASTILLO

DEPARTAMENTO DE PSICOLOGÍA EXPERIMENTAL

UNIVERSIDAD DE GRANADA

NOVIEMBRE 2013

Editor: Editorial de la Universidad de Granada
Autor: María Tortosa Molina
D.L.: GR 869-2014
ISBN: 978-84-9028-925-9

UNIVERSIDAD DE GRANADA

Universidad de Granada
Departamento de Psicología Experimental
Campus Universitario de Cartuja, s/n
Telf. +34-958243763 –Fax: +34 -958246239
18071 – Granada (España)

THE ROLE OF RACE AND EMOTIONAL EXPRESSIONS ON TRUST DECISIONS:

Tesis Doctoral presentada por **María Tortosa Molina** en el *Departamento de Psicología Experimental* para aspirar al grado de Doctora en Psicología, en el Programa de *Doctorado en Psicología* de la Universidad de Granada. En este trabajo se han respetado las pautas que establece la normativa de la Universidad de Granada para la obtención del título de Doctorado Internacional.

La tesis ha sido realizada bajo la dirección del los profesores María Ruz Cámara y Juan Lupiáñez Castillo, quienes avalan la calidad de la misma, así como la formación de la doctoranda para aspirar al grado de doctor.

Firmado en Granada, a 18 de Noviembre de 2013

La doctoranda: Los directores de tesis:

Fdo. María Tortosa Molina Fdo.: María Ruz Cámara Fdo.: Juan Lupiáñez Castillo

CONTENIDOS

RESUMEN GENERAL

La toma de decisiones es algo presente en nuestra vida diaria. En muchas de estas decisiones están implicadas las personas con la que interaccionamos, y se basan en la confianza y en las expectativas que tenemos acerca de los demás, considerando cuál será su reacción o su próximo movimiento. Para prever esto nos guiamos de diferentes factores; desde los rasgos físicos hasta el conocimiento personal que tenemos de la gente, la categoría social a la que pertenecen o incluso la información que nos proporcionan los otros sobre esa gente. Con todo esto tratamos de predecir el comportamiento de los demás.

En esta tesis doctoral estudiamos cómo influyen en situaciones de toma de decisiones de cooperación, ciertos factores sociales que extraemos a partir de información presente en los rostros de las personas con quien interactuamos, como son la raza y la expresión emocional. A través de una serie de estudios exploramos cómo estos factores modulan y sesgan nuestras interacciones sociales empleando un juego económico denominado Juego de Confianza.

Para ello hemos llevado a cabo una serie de experimentos en los que hemos recogido datos comportamentales, y también electrofisiológicos mediante el registro de electroencefalograma (EEG) y los potenciales evocados. En el juego de Confianza, los participantes se enfrentan a diferentes parejas de juego, mediante una simulación por ordenador que les muestra sus caras, y tienen que decidir si comparten parte de su dinero o no con ellos. Dicha pareja, a su vez, decidirá si es recíproca en la inversión, o por el contrario se queda con todo lo recibido.

En la primera serie experimental comenzamos estudiando el efecto a nivel comportamental de la emoción estableciendo diferentes contingencias entre dicha emoción y las tasas de reciprocidad de las parejas de juego; observamos un efecto de la expresión emocional de manera que los participantes cooperan más con aquellos que muestran expresiones de alegría que de enfado incluso después de repetida experiencia en la que la emoción expresada por la pareja de juego no predice su reciprocidad. Cuando las contingencias del juego entran en conflicto con las tendencias naturales (i.e., expresiones de enfado predicen mayor cooperación que expresiones de alegría), los participantes aprenden estas asociaciones y adaptan sus respuestas, aunque el aprendizaje es más lento que cuando las contingencias se adecuan a las consecuencias naturales (i.e., rostros sonrientes predicen mayor tasa de cooperación que rostros de enfado).

En una segunda serie, llevamos a cabo tres experimentos empleando el registro electroencefalográfico (EEG) e introdujimos la raza como otra variable de las parejas de juego a considerar además de su expresión emocional. Queríamos observar los potenciales evocados asociados a la emoción y la raza del rostro, y comprobar si se daban interacciones a lo largo del procesamiento. En este caso, ni la raza ni la emoción eran predictivas de la decisión de la pareja de juego, de manera que permitía observar cómo los participantes hacían uso de ellas a pesar de no contener ningún tipo de información respecto al futuro comportamiento del compañero. A nivel comportamental, en ausencia de otra información en la cual poder basar las decisiones, los participantes usaban las expresiones emocionales para guiar sus decisiones, sin embargo no ocurrió así en el caso de la raza, a pesar de un claro sesgo negativo en asociaciones implícitas para miembros de la raza negra, que abordamos mediante un Test de Asociación Implícita. Al igual que en el experimento de la serie anterior tendían a cooperar más con la gente que muestra una expresión de alegría que de enfado, incluso después de repetidas interacciones en ausencia de contingencia alguna entre expresiones emocionales y cooperación de la pareja de juego. En este caso analizamos tanto tasas de cooperación de los participantes como el registro de EEG mediante los potenciales evocados asociados a la cara (tomando como variables raza y emoción). Estos resultados mostraron una interacción en los potenciales N170 y P200, en zonas occipito-temporales, entre la raza y la emoción del participante que sugiere un procesamiento interrelacionado de aspectos variantes e invariantes del rostro.

Por otro lado, un tercer bloque de esta tesis, se centra en cómo las personas procesamos el feedback que recibimos de los demás después de haber tomado una decisión de cooperación o no cooperación. Basándonos en los estudios previos de un índice neural del procesamiento del *feedback* como la negatividad asociada al feedback (FRN, del inglés *feedback-related negativity*), queríamos investigar cómo afectan estas claves sociales de raza y emoción al procesamiento del feedback recibido. Este marcador se modula con las expectativas, de manera que ante un resultado que es peor de lo esperado se obtienen mayores amplitudes en la FRN que ante un resultado positivo o de ganancia. Predecíamos modulaciones en el FRN dependiendo de la emoción de la persona de la que procediese el feedback y de su raza. Hemos encontrado modulaciones consistentes en la FRN por la emoción, que serán discutidas a la luz de la teoría dopaminérgica (RL-*theory*, del inglés *reinforcement learning-theory*) y el procesamiento automático de la expresión emocional. En un último experimento de esta serie, incluimos la identidad de la persona como clave predoctora de su comportamiento, con un 70% de probabilidad (e.i., parejas cooperativas vs. no cooperativas). Queríamos comprobar si ante esta circunstancia, la emoción seguía influyendo tanto en el comportamiento como en el procesamiento del feedback. En este caso fue la identidad de la persona y no la emoción lo que interaccionó con el feedback, así como la que mostró mayores efectos en las tasas de cooperación. Aún así, la emoción siguió ejerciendo un efecto residual tanto a nivel comportamental en las tasas de cooperación como sobre la FRN.

En la discusión final de la tesis intentamos integrar los resultados obtenidos bajo el marco teórico de la toma de decisiones en las relaciones interpersonales, teorías evolutivas de la emoción y la automaticidad del procesamiento de las claves sociales.

CAPÍTULO 1

INTRODUCCIÓN

La cognición social es un área amplia que integra diferentes niveles de análisis. A nivel *social,* estudia la habilidad de evaluar, anticipar y reaccionar a los eventos sociales y cómo éstos influencian el comportamiento y la experiencia de las personas; a nivel *cognitivo*, estudia los mecanismos de procesamiento de la información que subyacen a las conductas sociales. En las últimas décadas, con los avances en neuroimagen, la neurociencia ha hecho su incursión en este campo, dando lugar a la neurociencia cognitiva social, y ha incorporado el nivel *cerebral*, que se centra en las redes neurales específicas que median estos procesos (Ochsner & Lieberman, 2001). Esta disciplina incipiente aborda temáticas centrales de contenido social como son actitudes y estereotipos, prejuicio, teoría de la mente, interacciones sociales, emoción, toma de decisiones, etc.

En esta tesis doctoral hemos estudiado las interacciones con los demás en la toma de decisiones en el ámbito de la neurociencia cognitiva social, a través de paradigmas de interacción social en los que hay que decidir si cooperar o no con alguien, e indagar en el uso que hacemos de la información que nos dan claves faciales como pueden ser la raza de la persona o su expresión emocional. A lo largo de la introducción expondremos estas temáticas desde la aproximación de la neurociencia cognitiva social. Comenzaremos con la habilidad de extraer información acerca de los estados mentales de los demás conocida como Teoría de la Mente, la cual nos permite acceder a la intención momentánea de la gente a través de su expresión emocional u otros aspectos variantes como la dirección de su mirada. Dado que el estímulo social más relevante en el que nos basamos para ello es el rostro de las personas, haremos una revisión de la bibliografía referente al procesamiento de las caras de los demás, qué aspectos son importantes y las áreas cerebrales implicadas en estos procesos. Nos centraremos también en el estudio de aspectos invariantes del rostro como la raza, mediante los cuales podemos acceder a una tendencia de acción más general del individuo, mediante procesos de categorización social.

Dado el interés específico de la presente tesis, haremos hincapié en la influencia que estos aspectos tienen en las interacciones sociales en general, y en particular en la toma de decisiones referente a confiar o no en los otros. Para ello revisaremos primero la investigación en el procesamiento de estas claves sociales, sus funciones a nivel social (como claves predictivas del comportamiento de los demás) y sus correlatos neurales a nivel electrofisiológico y de neuroimagen. Seguidamente introduciremos la toma de decisiones dentro del ámbito de la neuroeconomía y la teoría del juego, revisando algunas de las investigaciones que ponen de manifiesto correlatos psicológicos y neurales en la toma de decisiones y expondremos cómo hemos implementado los factores de raza y emoción como objeto de estudio en los juegos de confianza, a la hora de decidir cooperar o no con los demás.

1. *Marco teórico general: Neurociencia Social*

> "…as members of an intensively social, cooperative, and competitive species, our ancestors' lives depended on how well they could infer what was on one another's minds…"
>
> (Baron-Cohen, Mindblindness)

Dada la naturaleza social de los seres humanos, la cognición social es algo ubicuo. Al interaccionar con alguien entra en juego no sólo nuestro conocimiento perceptivo del mundo, sino del contexto social, y la forma como categorizamos socialmente a los demás. Además, entra en juego el conocimiento que tenemos de nuestros estados mentales, de los estados mentales de los demás, de cómo pensamos que los demás perciben nuestros estados mentales, de cómo los demás piensan que nosotros percibimos sus estados mentales, y de cómo pensamos que los demás piensan que nosotros percibimos sus estados mentales.

Muchos de estos procesos son automáticos, en cuyo caso son muy eficientes; es decir, son los estímulos externos o situaciones, los que sin nuestro conocimiento o conciencia controlan nuestros estados internos. Hay estudios que señalan la automaticidad en la cognición social y sus efectos sobre las decisiones personales y el comportamiento (Bargh & Ferguson, 2000; Taylor & Fiske, 1978; Cañadas, Rodríguez-Bailón, Milliken, & Lupiáñez, (2013). Esto no excluye que a la vez seamos capaces de emplear control para comportarnos de manera estratégica y deliberada. Un caso extremo de comportamiento social controlado y anómalo sería el que se da en autismo, resultando en una conducta poco eficiente a nivel social.

Las impresiones que nos formamos de los demás ocurren muchas veces de esta manera; el sexo de la gente, su raza, las emociones que expresan con su rostro, la reputación o la edad son atributos que influyen en nuestro comportamiento y en las relaciones que establecemos con ellas. Podemos categorizar y aprender de los demás como "grupo" o "tipo de persona", de forma que podemos actuar igualmente ante alguien que no conocemos, aplicando nuestro conocimiento previo acerca de su grupo de pertenencia, que a su vez conlleva la activación de estereotipos y prejuicios (Dovidio, Evans, & Tyler, 1986; Frith & Frith, 2012). En este sentido, los aspectos invariantes de las características físicas de los demás (sexo, edad[1], raza) permiten llevar a cabo una categorización social y pueden activar automáticamente estereotipos hacia las personas de esas categorías (Bargh & Williams, 2006). Estas influencias automáticas son importantes en las interacciones sociales para la autorregulación y las decisiones que tomamos en nuestra vida diaria cuando interactuamos con los demás. Por ejemplo, como señalan Frith y Frith (2012), "aprender sobre el grupo de procedencia y reputación de los demás es crucial para las interacciones sociales que se basan en la confianza", sobre todo en ausencia de conocimiento previo. Así, en un primer encuentro con alguien el rostro ofrece muchas características que permiten la formación de categorías sociales, y nos basamos en esas claves (i.e., el color de la piel, la edad, el sexo) para formarnos una primera impresión y hacer inferencias sobre su conducta, lo que a su vez influirá en nuestros juicios de confianza. Pero también aprendemos sobre los demás como individuos, esto es, de manera individualizada, (véase Brewer, 1988 y Fiske & Neuberg, 1990[2]) por las características específicas que conocemos de ellos (i.e., su temperamento y carácter) lo que nos cuentan, etc.

Por otro lado, en la interacción directa cara a cara podemos usar otras claves variantes del rostro para interpretar estados o intenciones momentáneas de los demás. Aspectos como la dirección de la mirada de alguien y su expresión emocional nos permiten hacer inferencias acerca de su estado mental, y predicciones sobre sus intenciones y acciones inmediatas. Esta capacidad humana es posible porque nuestro cerebro es un "cerebro social" (Brothers, 1990), que ha evolucionado para permitirnos llevar a cabo conductas de tipo social, desarrollando capacidades como son la teoría de mente, la empatía y la toma de decisiones.

Así, como hemos mencionado anteriormente, gran parte del conocimiento que tenemos del mundo se basa en el conocimiento de la mente de los demás y el conocimiento de nuestra propia mente. Esto requiere que seamos capaces de distinguir entre el conocimiento de "fuera" que tenemos sobre el mundo externo (no social –e.g. hay nieve fuera, estamos en invierno), el de la mente de los demás (e.g. busca su abrigo, tiene frío) y el de nuestra propia mente (siento el frío, tirito). Esto se conoce como Teoría de la mente (ToM, del ingles *Theory of Mind*; Leslie, 1987): la atribución de estados mentales (un yo *pensante* y *sintiente*) a los demás y a uno mismo, y de la representación que los demás tienen de nuestro yo. Permite hacer predicciones e inferencias acerca de los estados internos de otras personas para predecir su conducta (Premack & Woodruff,

[1] Entiéndase invariante para este caso en el curso temporal en que transcurre la interacción.

[2] Estos autores plantean un modelo de formación de impresiones y percepción personal desde la cognición social en el que proponen dos procesos básicos: la categorización y la individualización.

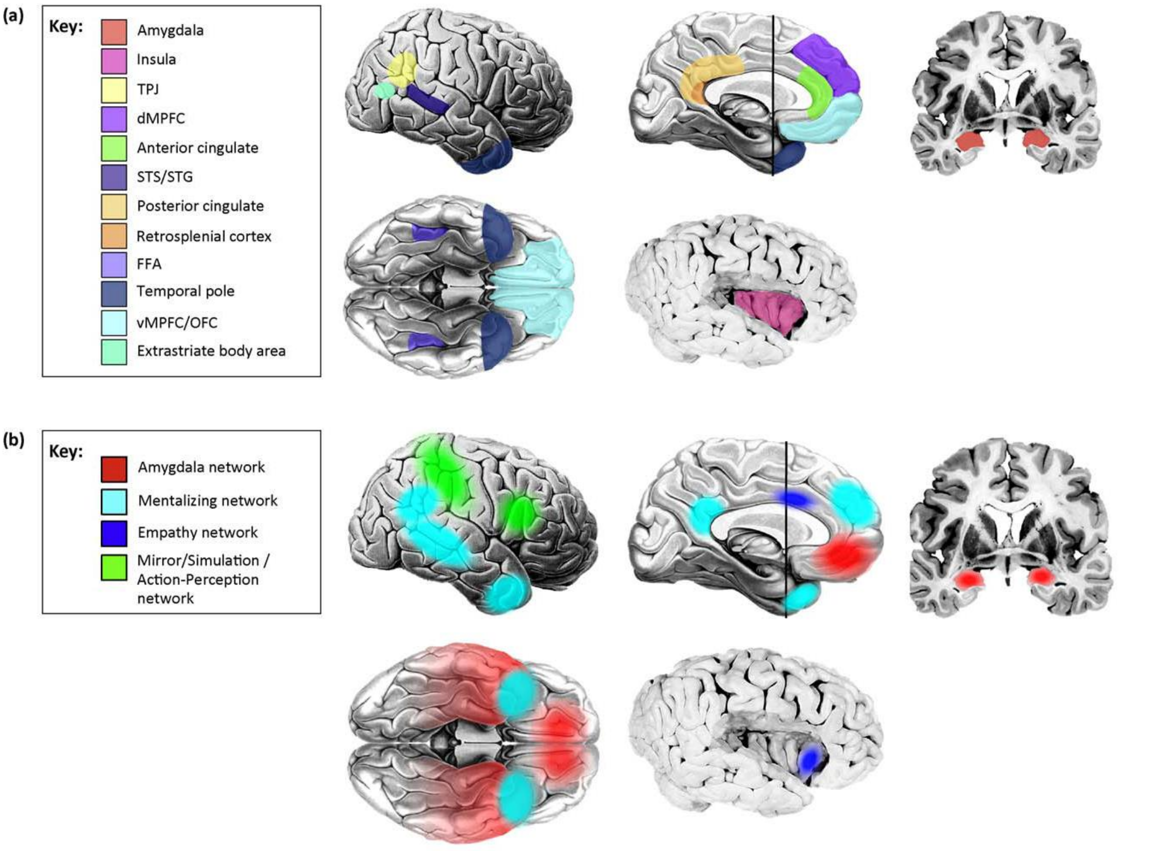
1978), tener una explicación de su comportamiento y atribuir intenciones, basadas en creencias, deseos y pensamientos (Baron-Cohen, 1997). Parte de estas inferencias están basadas en la influencia que las claves sociales de los demás ejercen en nuestro comportamiento e interacciones con ellos.

Una de las manifestaciones más tempranas de ToM es la capacidad de fingir, que surge en los niños en el tercer año de vida (Leslie, 1987; e.g. jugar con una taza vacía pretendiendo beber de ella). Algo más tarde, a los 3-4 años, empiezan de forma explícita a atribuir estados mentales –creencias, pensamientos, conocimientos- a los demás para explicar sus acciones (Saxe, Carey, & Kanwisher, 2004). Las tareas sobre teoría de la mente se basan en hacer razonar sobre las creencias e intenciones de otros, en concreto la capacidad para atribuir creencias falsas. La tarea clásica que se usa con niños de edades tempranas para ver si han desarrollado esta capacidad es la *tarea de la falsa creencia*, (Wimmer & Perner, 1983). En esta tarea, conocida como la tarea de Sally y Anne, se presenta al niño una situación (suele hacerse con viñetas ilustrativas) en la que Sally, deja su muñeca en su cesta y abandona la escena. En su ausencia, Anne coge el objeto y lo cambia de sitio, poniéndolo en una caja. ¿Dónde buscará Sally la muñeca al volver? Los niños menores de 4 años suelen fallar esperando que Sally busque la muñeca en la caja, ya que el niño entiende que es su propio estado mental y no el estado del mundo (Sally ha salido de la habitación y no ha visto que la muñeca ha cambiado a la caja) lo que causa la acción.

La empatía es otra manifestación de la teoría de la mente. Es la habilidad para experimentar los estados emocionales de otros (Carrington & Bailey, 2009) y está relacionada con la atribución de estados mentales. Entender las emociones requiere una atribución causal sobre las intenciones que subyacen a una acción (Olsson & Ochsner, 2008); difícilmente se podría entender la alegría manifestada por alguien sin atribuir a esa persona pensamientos, creencias o deseos. Algunos autores distinguen entre dos subprocesos: uno se correspondería a las reacciones más emocionales de la empatía y otro a las cognitivas (Shamay-Tsoory, 2011). Las primeras implican reacciones afectivas ante las experiencias observadas de otro, que suelen darse por contagio y reconocimiento emocional. Esto es lo que ocurre de manera más temprana a nivel filogenético (e.g. el contagio del llanto en bebés; véase De Waal, 2008) y tiene que ver con el sistema de neuronas espejo. Por otro lado, la empatía cognitiva hace referencia a entender la perspectiva del otro, poniéndose en su punto de vista, lo cual viene siendo lo que hemos definido como la ToM. De hecho, estudios en empatía han puesto de manifiesto que se activan áreas cerebrales de activación relacionadas con teoría de la mente (Shamay-Tsoory, 2011; Vollm, Taylor, Richardson, Corcoran, Stirling, et al., 2006), tales como la ínsula anterior y la corteza cingulada anterior (Singer, Seymour, O'Doherty, Kaube, Dolan, & Frith, 2004; Singer, Seymour, O'Doherty, Stephan, Dolan, & Frith, 2006).

La ToM y la capacidad de empatizar están a la base de la influencia y las reacciones afectivas que provocan en nosotros las claves sociales presentes en la cara de los demás. Respecto a las áreas cerebrales relacionadas con ToM y empatía, muchos estudios durante los últimos años han empleado técnicas de neuroimagen para abordar las redes cerebrales implicadas, así como los diferentes procesos cognitivos a los que contribuyen las diferentes áreas. Queda fuera del alcance teórico de esta introducción hacer una amplia revisión de las regiones cerebrales que han sido implicadas en teoría de la mente a lo largo de la literatura, dada la cantidad de estudios de neuroimagen que han puesto su atención en ello (véase Saxe, 2006; Gallagher & Frith, 2003; Carrington & Bailey, 2009). Sólo haremos mención de un mapa cerebral general de las bases neurales que consistentemente parecen intervenir en la capacidad de mentalizar.

Entonces, ¿qué estructuras neurales subyacen a esta habilidad tan compleja? Aunque no hay que perder de vista la variabilidad metodológica que supone abordar un constructo tan amplio y subjetivo como es el hacer inferencias sobre los estados mentales de los demás, se han identificado a pesar de eso regiones fundamentales que ofrecen patrones de actividad diferenciales en tareas que implican mentalizar. Se trata de una amplia red que incluye las áreas de la imagen presentada más abajo (Figura 1). En el lóbulo prefrontal, por ejemplo la corteza prefrontal medial (CPFm) se ve implicada en mentalizar; la corteza orbitofrontal (COF) interviene en regular la conducta social; el giro frontal inferior (GFI), se ve envuelto en simulación e imitación, y en empatía emocional (Shamay-Tsoory, 2011). Dentro del lóbulo temporal, la unión temporoparietal (TPJ, del inglés *temporo-parietal junction*; Gallagher & Frith, 2009; Saxe & Wexler, 2005) se ve activa en tareas de ToM; los polos temporales almacenan recuerdos personales semánticos y episódicos, y se relacionan con empatía (Vollm et al., 2006). Según el estado mental en concreto y el paradigama, si se trata de historia en las que interviene el lenguaje y la memoria (Calarge, Andreasen, & O'Leary, 2003; o ilustraciones animadas (e.g. evocando una secuencia de eventos con tres posibles finales; Brunet, Sarfati, Hardy-Bayle, & Decety, 2000), estas redes se configurarán de un modo u otro, y en conjunción con estructuras subcorticales asociadas al sistema límbico como amígdala y la ínsula.



TRENDS in Cognitive Sciences

Figura 1. Representación esquemática de las regiones cerebrales asociadas con la atribución de estados mentales: creencias, deseo, percepciones y/o emociones (de Kennedy y Adolphs, 2012).

Gallotti y Frith (2013) han propuesto recientemente una aproximación teórica distinta. Considerando que las teorías clásicas de teoría de la mente son reduccionistas e individualistas, exponen una visión teórica más amplia e integradora, desde el interaccionismo y el “*we-mode*”; entender a los otros y sus acciones como parte de un hacer conjunto, como dirigidas a un objetivo perseguido en común. Desde esta perspectiva, dado que la actividad cognitiva consiste en hacer que las relaciones entre un organismo y su ambiente tengan sentido, cuando el ambiente es social este hacer sentido tiene que ocurrir de una manera conjunta-participativa. En esta dirección, una nueva aproximación metodológica y conceptual que se viene desarrollando en los últimos años aborda de manera más amplia la investigación relacionada con interacción social. Se trata del registro de actividad cerebral de dos personas a la vez mediante estudios con electroencefalograma (EEG) o imagen por resonancia magnética funcional (IRMf), mientras realizan una tarea en la que tienen que interactuar. Por ejemplo estudios con EEG dual o con *hiperscanning* (Yun, Watanabe, & Shimojo, 2012; véase también Hari, Himberg, Nummenmaa, Hämäläinen, & Parkkonen, 2013, para una revisión) han hallado correlatos de sincronización entre dos personas recogiendo actividad cerebral de ambos al mismo tiempo. Citaremos alguna de estas investigaciones relacionadas con toma de decisiones en el último apartado de la introducción, ya que son relevantes para una aproximación más ecológica a la toma de decisiones dentro de la neurociencia cognitiva social

Los estudios presentados en este trabajo de investigación se dirigirán a estudiar procesos de toma de decisiones basándonos en el rostro de una persona como principal fuente de información. Por ello, en el siguiente apartado revisaremos las claves sociales presentes en el rostro que pueden ser reguladores en nuestras interacciones con los demás y de las cuales extraemos información, así como las propuestas teóricas o modelos que tratan de explicar el procesamiento de caras y los circuitos cerebrales implicados.

2. *Percibiendo rostros*

Como se ha expuesto, poseemos habilidades sociales que nos permiten extraer información sobre los estados mentales de los demás. La primera información disponible en las relaciones interpersonales es la cara de la otra persona, que es la clave más distintiva para identificar a una persona (Bruce & Young, 1986). Pero además es el primer atisbo comunicador al encontrarnos con alguien. En ausencia de conocimiento previo, el rostro pone de manifiesto gran cantidad de información referente a su edad, sexo, raza, estado emocional, y deja entrever gran variedad de significados; se pueden extraer rasgos de personalidad (Berry & McArthur 1985), atractivo (Wilson & Eckel, 2006), estatus y dominancia, (Keating, Mazur, & Segall, 1977), dirección de la atención visual (Bayliss & Tipper, (2005), intenciones y motivaciones, Horstmann, (2003). Algunas interpretaciones son bastante directas como el sexo, mientras que otras son más subjetivas y susceptibles de error, como el carácter de una persona, su pertenencia a un grupo de alto status o su orientación sexual (Tskhay & Rule, 2013).

De alguna manera, toda esta información modula el comportamiento, muchas veces de un modo automático, en la forma de interactuar y tomar decisiones con los demás. Además guían nuestras primeras impresiones de la gente (Bar, Neta, & Linz, 2006), mejoran la comunicación ya que pueden matizar y realzar la comprensión del habla, y proporcionan información acerca de los estados mentales de los demás. Ante situaciones o estímulos ambiguos del ambiente que no sabemos cómo interpretar, podemos hacer uso de las reacciones faciales de los otros para clarificar o desambiguar la situación.

Por lo tanto, los marcos teóricos del procesamiento de caras, que expondremos a continuación, partieron de dos preguntas; 1- dada su función principal de identificación, ¿hay algo especial acerca de las caras para que se identifiquen como tal y de manera diferenciada al resto de estímulos?; 2- dado su papel fundamental en la comunicación, ¿existen mecanismos de base que diferencien el identificar a alguien (“*es mi vecino”*) de lo que nos comunica (“*está preocupado”*)? Presentaremos a continuación estos modelos que tratan de explicar el procesamiento de caras para después pasar a lo que sabemos acerca de cómo el cerebro extrae la información de las caras, y los correlatos neurales relacionados.

2.1. Modelos teóricos del procesamiento de caras

En la década de los 80 autores como Ellis, H., Bruce y Young proponían que hay un sistema de procesamiento específico para las caras. Esta idea se materializó a partir de dos hallazgos científicos relevantes: 1) datos neuropsicológicos de pacientes con prosopagnosia, un síndrome asociado a lesiones en la corteza ventral occipito-temporal donde se ve deteriorada la habilidad para el reconocimiento de caras (Damasio, Damasio, & Van Hoesen, 1982; McNeil & Warrington, 1993; De Renzi, 1997); 2) estudios con primates no humanos, en los que registros unicelulares identificaron neuronas en el STS y la corteza temporal inferior que responden selectivamente a caras (Perrett, Mistlin, & Chitty 1987).

El hecho de que se haya disociado el procesamiento de caras del procesamiento de otros objetos llevó al desarrollo de modelos explicativos específicos al procesamiento de caras. En estos modelos cognitivos, los autores proponen rutas de procesamiento funcional y neuroanatómicas separadas para identidad y procesamiento de aspectos variables de las caras, desde la representación visual hasta su interpretación. Bruce y Young (1986) propusieron un marco teórico en el que describen rutas separadas de procesamiento para identidad, donde se procesa la información fisionómica, y otras para el reconocimiento de expresión o para movimientos relacionados con el habla. Según este modelo, basado principalmente en disociaciones con pacientes con lesión cerebral, los procesos son funcionalmente independientes, por ejemplo, la expresión emocional se procesaría de forma independiente y no influiría en el procesamiento de la información de la identidad.

Hoffman y Haxby (2000), algunos años después y bajo la luz de los primeros estudios de neuroimagen, localizan anatómicamente los componentes del procesamiento de caras en regiones neurales específicas. Diferenciaron entre un sistema central (*core system*), en la corteza visual occipito-temporal extraestriada, y otro más extenso. El primero se encargaría de representar tanto la identidad como los aspectos variables de las caras, a través de diferentes rutas de procesamiento. Este sistema central despliega y establece conexiones con un sistema amplio, relacionado con aspectos atencionales, emocionales, auditivos. Además, el reconocimiento de la identidad estaría anatómicamente disociado de la percepción de expresión facial, detección de la mirada y aspectos del habla, aspectos estos últimos asociados a la comunicación social.

Una aproximación alternativa a los modelos anteriores que se basan en el contenido informativo (identidad vs. expresión), fue propuesta por Calder y Young (2005). Según estos autores, el nivel de diferenciación funcional y anatómico no es tan claro. Tras una revisión a la literatura reciente, concluyen que la investigación en prosopagnosia y déficits en reconocimiento de la emoción no corroboran una doble disociación tan evidente. Así como tampoco está claro a qué nivel de análisis se bifurcan las rutas de la identidad facial y de la expresión facial. Proponen un procesamiento basado en las diferentes propiedades físicas y demandas de procesamiento de la información. Aunque reconocen cierto grado de separación neural entre las rutas de procesamiento de la identidad y la expresión, atribuyen al STS un rol más integrador, que recibiría información de otras modalidades sensoriales y la integraría.

En los últimos años, Atkinson y Adolphs (2011) proponen que los sistemas neurales del procesamiento de caras son más interactivos y flexibles de lo que hasta entonces se planteaba, haciendo hincapié en el papel del área occipital inferior (OFA, del inglés *occipital face area*). Implicada en la percepción inicial de claves visuales como los rasgos de una cara o sus formas en etapas tempranas del procesamiento visual, según el modelo de Hoffman y Haxby (2000), para estos autores tendría un rol más interactivo con otras regiones de procesamiento, dependiendo de la naturaleza del estímulo y la tarea. Así, la OFA se ha visto implicada en juicios de confianza (Dzhelyova, Ellison, & Atkinson, 2011) o en integración de información entre identidad y expresión (Gothard, Battaglia, Erickson, Spitler, & Amaral, 2007). Así, los dos sistemas paralelos (central y extendido) podrían estar interactuando a lo largo de las fases del procesamiento.

Desde otra perspectiva teórica, se ha diferenciado entre aspectos invariantes y variantes en las caras, (Hoffman & Haxby, 2000; Cloutier, Turk, & Neil Macrae, 2008). Una diferencia a la hora del procesamiento de unos y otros es que mientras los aspectos variantes requieren de monitorización constante durante las interacciones sociales, la identidad se procesa en un primer momento y no requiere de esta continua actualización. Entre los aspectos dinámicos estarían las expresiones emocionales, movimientos de los ojos y movimientos articulatorios del habla, que tienen que ver con la comunicación social y la ToM, ya que de ellos inferimos el estado actual de una persona, su intención momentánea.

Por último, haremos una distinción funcional entre codificación *configural* (holística) y analítica (de rasgo; Maurer, Le Grand, & Mondloch, 2002). El procesamiento *configur*al codifica las relaciones espaciales entre rasgos (relaciones de segundo orden). Se trata de un procesamiento holístico, en el que se basa el reconocimiento de la identidad. Tendemos a procesar las caras como una todo (de manera *gestáltica*), aunque hay estudios que han encontrado diferencias culturales. Mediante técnicas de movimientos oculares, se ha encontrado que en los asiáticos se hace más patente el procesamiento holístico, ya que realizan más fijaciones en el contorno externo, respecto a los occidentales, que se fijan más en la zona de los ojos y la boca (Blais, Jack, Scheepers, Fiset, & Caldara, 2008). El procesamiento analítico, se refiere a un procesamiento rasgo a rasgo de los elementos constituyentes de la cara (ojos, boca, nariz…). Evolutivamente, el procesamiento *configural* se desarrolla de manera más lenta que el procesamiento de los rasgos, que es el primero en madurar en los niños, (Cohen & Cashon, 2001). Por otro lado, también hay estudios que muestran que caras de la propia raza se procesan de manera más holística que caras de otra raza (Michel, Rossion, Han, Cheng, & Caldara, 2006).

La influencia de algunas claves variantes e invariantes en la toma de decisiones es de especial interés para el desarrollo de la presente tesis, y nos centraremos en ellos en apartados subsiguientes. Pero antes vamos a describir brevemente las redes cerebrales y correlatos neurales asociados al procesamiento de caras.

2.2. Correlatos neurales asociados al procesamiento de caras.

La incursión de las técnicas de neuroimagen y electroencefalografía ha permitido empezar a vislumbrar correlatos neurales de cómo procesa el cerebro los rostros. Apoyada por la neuroimagen, gran cantidad de investigación en neurociencia cognitiva durante las últimas décadas apunta a la existencia de redes neurales especializadas para el procesamiento de las caras (McCarthy, Puce, Gore, & Allison, 1997; Haxby, Hoffman, & Gobbini, 2000; 2002; Hoffman & Haxby, 2000; Ishai, Schmidt, & Boesiger, 2005; Kanwisher, 2000; Kanwisher, McDermott, & Chun, 1997; pero véase Gauthier &

Logothetis, 2000); aunque este sigue siendo un tema controvertido.

Las redes cerebrales identificadas se extienden en regiones occipito-temporales, sobre todo para el reconocimiento de identidad, aspectos variables e invariables del rostro. De acuerdo al modelo de Hoffman y Haxby, el sistema central se correspondería con zonas en giro occipital inferior (GOI), el cual manda señales al giro fusiforme lateral (GFL) y al surco temporal superior (STS). Dentro de estas redes, las zonas más relacionadas con aspectos invariantes para el reconocimiento de identidad, parecen yacer en el giro fusiforme lateral, donde se ha ubicado el área fusiforme de las caras (FFA; ver más abajo). Este despliega y establece conexiones con diferentes redes implicadas en comunicación social más variantes, en las que intervienen desde zonas visuales de corteza occipital, hasta el sistema límbico, asociado a procesamiento emocional, regiones parietales de atención espacial que se activan ante la detección de dirección de la mirada, junto con el surco temporal superior, áreas auditivas relacionadas con el procesamiento de los sonidos del habla y la lectura de labios (Haxby et al., 2000).

Algunos autores, partiendo del hallazgo de una región más activa en la corteza occipito-temporal, el giro fusiforme derecho, cuando se ven caras que cuando se ven otros objetos, muestran evidencias de esta especificidad neural para caras. Kanwisher (1997), propuso que la existencia del FFA era una evidencia de la especificidad de dominio para el procesamiento de caras (Kanwisher et al., 1997). Sugieren que al reconocimiento de caras y de objetos subyacen procesos implementados en distintas áreas cerebrales, y defienden que el FFA está específicamente involucrado en la percepción de caras.

En contraposición a la perspectiva del procesamiento específico para caras, Gauthier y cols. (1999, 2000) proponen la hipótesis de la experiencia de nivel subordinado (*subordinate level expertise*; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Gauthier, Skudlarski, Gore, & Anderson, 2000); mientras que el resto de los objetos los solemos reconocer a un nivel básico (e.g. coches, pájaros) procesamos las caras no como una categoría básica sino a un nivel individual (es decir, Lucas, Ana, mi primo). El nivel de categorización al que procesamos los estímulos que a su vez viene determinado por la experiencia-pericia es crítico para explicar estos diferentes niveles de procesamiento. De manera que cuando se adquiere experiencia discriminando entre miembros homogéneos de una clase, el proceso de reconocimiento y los sustratos neurales correspondientes serán los mismos que los usados para el reconocimiento de caras (Rossion, Gauthier, Tarr, Despland, Bruyer, Linotte, & Crommelinck, 2000). Mantienen que el FFA interviene en el reconocimiento preciso de categorías de objetos bien conocidos, con los que hemos tenido mucha experiencia en nuestra vida, aplicable tanto a caras como a otras categorías.

Así, en múltiples estudios con IRMf, se muestra activación en FFA por los efectos de la experiencia-pericia en la materia; por ejemplo, expertos en las categorías de coches y pájaros mostraron una activación similar en la zona del FFA a la provocada por caras (Gauthier et al., 2000). Esta línea de investigación sugiere que una considerable parte de la diferencia en activación típicamente encontrada entre caras y objetos se explica por el nivel de categorización al que procesamos los estímulos, que a su vez viene determinado por la experiencia-pericia. Dado que tenemos mayor experiencia con caras que con cualquier otra categoría de estímulos en el mundo, las caras tienen mayor nivel específico de categorización que otros objetos, y sería esto lo que las haría especiales y no tanto por un componente de la categoría de objeto *per se*.

Además de la evidencia de circuitos neuroanatómicos y áreas específicas especializadas, vislumbrada a partir de estudios con pacientes neuropsicológicos y técnicas de neuroimagen en los últimos 25 años, medidas electrofisiológicas como son los potenciales evocados (ERPs, del inglés *event-related potentials*) han detectado consistentemente algunos correlatos neurales indirectos asociados al procesamiento de los rostros. Ante la presentación de caras junto con otros estímulos, se desencadenan ERPs diferenciales que son más sensibles a las caras, y se encuentra actividad diferencial en algunas zonas del cerebro. En concreto, el potencial N170 es un componente negativo que ocurre en localizaciones occipito-temporales alrededor de los 170 ms después de la aparición de una cara y es más pronunciado para caras que para otro tipo de estímulos, como casas, coches o manos (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Rossion et al., 2000; Eimer, 2000). Junto a datos neuropsicológicos y registros unicelulares, esto sugiere la existencia de neuronas especializadas cuya actividad está en sintonía con la detección de rostros humanos. El N170 podría reflejar parte de este mecanismo neural que, derivado de regiones occipito-temporales, coincide con la localización del área fusiforme (FFA; Kanwisher et al., 1997) región esta que muestra mayor activación ante caras en comparación con otros objetos.

Si bien el N170 no es una respuesta exclusiva para caras, sí apunta a un proceso específico para dicho estímulo; pruebas de ello son que su amplitud se incrementa y se retrasa su latencia ante caras invertidas en comparación a cualquier otra categoría de estímulos u objetos, que no presentan este efecto de inversión (Bentin et al., 1996; Rossion et al., 2000). Este hecho se explica por la pérdida de información *configural* o de relaciones espaciales entre las partes (Rossion, Delvenne, Debatisse, Goffaux, Bruyer, Crommelinck, & Guérit, 1999); también este incremento se ha explicado por la dificultad que conlleva el procesamiento de caras invertidas frente a no-invertidas (George, Evans, Fiori, Davidoff, & Renault, 1996; Yin, 1969), o porque se recluten además de las áreas específicas sensibles al procesamiento de caras, áreas más generales, implicadas en procesamiento de objetos (Haxby, Ungerleider, Clark, Schouten, Hoffman, & Martin, 1999).

Es además sensible a los rasgos intrínsecos de una cara (Bentin et al., 1996; Eimer, 2000). Así, cuando rasgos internos o externos de un rostro son eliminados, el N170 se ve atenuado, aunque sigue siendo más amplio en comparación a estímulos como coches o manos. El hecho de que se vea atenuado conforme el estímulo se aleja más de sus condiciones óptimas como cara, sugiere que el N170 está ligado a estadios tardíos del procesamiento estructural de caras (Eimer, 2000), que hacen categorizar una cara como tal (Bentin el at., 1996). Estudios posteriores muestran diferencias más tempranas en el procesamiento temporal de caras, indicando que comienza ya a los 100 ms (Herrmann et al., 2005). Por otro lado, no se puede descartar completamente la posibilidad de que estas diferencias en amplitud se deban a diferencias en propiedades visuales de bajo nivel entre caras y otros objetos (Rossion et al., 2000).

Concluyendo, dependiendo de cómo se aborde esta cuestión y de la metodología utilizada se pueden llegar a datos que apoyen una especificidad de dominio o un mecanismo común de procesamiento, quedando por tanto sin resolver lo que hay de especial acerca del procesamiento del rostro diferente al de otros estímulos.

En los próximos apartados nos centraremos en los aspectos de interés para los estudios experimentales de esta tesis, esto es, la raza como aspecto invariante de categorización social, y la expresión emocional como aspecto variante mediante el cual accedemos a estados mentales de los demás haciendo uso de la capacidad de mentalizar o ToM.

3. *Aspectos invariantes del rostro. La raza en las relaciones interpersonales.*

Como hemos mencionado en un apartado previo, hay aspectos físicos invariantes del rostro que permanecen estables de manera más o menos relativa a lo largo del tiempo, mediante los cuales identificamos y categorizamos a una persona. La edad, el sexo, y la raza son quizá los atributos invariantes más importantes y salientes (Bruce & Young, 1986), aunque recientemente se han estudiado la influencia de otros atributos invariantes más sutiles como el atractivo, la orientación sexual o la religión, (Tskhay & Rule, 2013).

Cuando tenemos que tomar decisiones acerca de si confiar o no en alguien, es en este tipo de claves sociales en las que nos basamos cuando no hay una historia personal previa con la persona. Si conocemos a esa persona, usamos la información individualizada que tenemos de ella; pero si se trata de un desconocido, lo categorizamos como perteneciente a una categoría social, en función de la información que extraemos de su físico, sobre todo, de su cara. Esto nos permite aplicar a esa persona el conocimiento y creencias que tenemos de su grupo de pertenencia (conocimiento basado en cierta medida en estereotipos; Stanley, Sokol-Hessner, Fareri, Perino, Delgado, Banaji, & Phelps, 2012). De entre estas características invariantes, nos vamos a centrar en el estudio y procesamiento de la raza y su influencia en las decisiones de cooperación, por ser uno de los factores importantes en las relaciones interpersonales, y en concreto en el ámbito de la cooperación, dadas las implicaciones respecto al prejuicio que conlleva en las relaciones intergrupales (Brewer, 1998), además de porque es el atributo invariante que manipulamos en nuestras series experimentales.

La influencia de la raza en la percepción y en el comportamiento ha sido estudiada desde hace décadas en el campo de la psicología social, así como la naturaleza y efectos del prejuicio. Desde la década de los 50 se llevan a cabo estudios de actitudes étnicas y respuesta emocional asociada (Rankin & Campbell, 1955; Westie & De Fleur, 1959), y en las últimas décadas se han hecho intentos de desarrollar medidas fiables para abordar estos constructos. Más recientemente, con la inclusión de técnicas electrofisiológicas y de neuroimagen (Phelps, O'Connor, Cunningham, Funayama, Gatenby, Gore, & Banaji, 2000; Golby, Gabrieli, Chiao, & Eberhardt, 2001; Ito & Bartholow, 2009), se han abordado los mecanismos cerebrales y los sustratos neurales subyacentes a la percepción de la raza. En este ámbito, la investigación reciente desde la neurociencia social busca dilucidar cómo la raza influencia el pensamiento y la acción (Eberhardt, 2005). Así, muchas son las preguntas que han surgido desde la cognición social, desde las más teóricas a las más aplicadas: ¿cómo se procesa la raza de una persona en el cerebro?, ¿por qué reconocemos mejor a gente de nuestra propia raza que de otra?, ¿cómo influye al comportamiento del que la percibe? Los siguientes subapartados tratarán de la automaticidad de este procesamiento y abordará también el tema del prejuicio.

3.1. La automaticidad de la raza

Ya en 1914 se demostró que las caras de la raza propia se reconocen mejor que las caras de otras razas (Feingold, 1914). Esto se conoce como "el efecto de la otra raza" (*other-race effect*) y las explicaciones que se han dado tienen que ver con los efectos de la experiencia, (Brigham & Barkowitz, 1978), de forma que el mayor conocimiento y experiencia con los miembros de nuestra propia raza a lo largo de nuestra vida modela plásticamente nuestro cerebro para fovorecer el procesamiento de los miembros de nuestra raza.

Efectivamente, la gente encuentra mucha similitud entre los miembros de otra raza (Caldara, Thut, Servoir, Michel, Bovet, & Renault, 2003). También al tratar de recordar, el reconocimiento para las caras de la propia raza es mejor que para las de otra raza (Golby et al., 2001), y se ha comprobado que esta discriminación mejora con la experiencia, conforme nos vamos haciendo más expertos con los miembros del exogrupo. También se han implicado en este efecto factores de categorización social (la raza como un rasgo visual, que se procesa a nivel holístico; Levin, 2000). Se piensa igualmente que puede deberse a que prestamos mayor atención a aquello con lo que tenemos más experiencia, en este caso, a los miembros de la propia raza, o a factores de tipo emocional; así, las respuestas emocionales ante miembros de la propia y de otra raza podrían afectar al procesamiento perceptivo (Golby et al., 2001).

A nivel cerebral, la percepción de la raza comienza cuando visualizamos las características físicas de la cara de una persona (Ito & Bartholow, 2009), ya que hay rasgos característicos como el color de la piel o la forma de las facciones (e.g. labios gruesos en el caso de la raza negra) que son discriminativos de una raza. Esto ocurre de una manera rápida y automática y comienza en regiones visuales primarias de la corteza occipital. De hecho, así lo evidencia los resultados en investigación con potenciales evocados, que muestra que la extracción de información referente a la raza comienza en los primeros 100 ms a partir de la presentación de la cara (Ito & Urland, 2003) lo que sugiere que de alguna manera en este momento ya codificamos información referente a categorías sociales, aunque sea a nivel de características preceptivas básicas. Esto es relevante, dado que una rápida percepción en base a rasgos salientes es crucial para la rápida categorización social de la persona percibida, y consecuentemente su interpretación, percepción final y nuestra interacción con ella. Todo ello se ve avalado por un diferente patrón de actividad neural en respuesta al propio grupo comparado con el exogrupo (Golby et al., 2001). Por ejemplo, una mayor respuesta del FFA para caras de la propia raza que de otra.

También la *amígdala* muestra mayor activación ante caras de otra raza que ante caras de la propia (Phelps et al., 2000); aunque esta activación parece estar más relacionada con las connotaciones afectivas en la que es interpretado el estímulo racial más que de sus características perceptivas distintivas. De hecho, en un estudio con resonancia magnética funcional se muestra cómo la activación en la amígdala se da sobre todo cuando se establece mirada directa (vs. desviada) con los miembros de la otra raza, lo que sugiere una connotación de potencial detección de amenaza en lo que se procesa (Richeson, Todd, Trawalter, & Baird, 2008), y no sólo la raza en sí. También la FFA parece tener alguna implicación en la identificación de la raza de otro individuo. Un estudio reciente encontró mayor activación en FFA ante caras de la propia raza en comparación con caras de otra raza (Golby et al., 2001), además de una mayor activación en FFA con la propia raza durante la codificación de caras para una tarea posterior de reconocimiento. Esto puede estar a la base del hecho de que las caras de la raza propia se reconozcan mejor que las caras de otras razas (“efecto de la otra raza”); así, las caras de otra raza eran caras menos familiares y más difíciles de reconocer que las de la propia raza, y se codifican a nivel categórico (homogeneidad perceptiva; Kubota, Banaji, & Phelps, 2012). También la corteza posterior cingulado interviene en aspectos de raza y muestra diferentes respuestas neurales ante endogrupo vs. exogrupo, aunque esta activación podría estar mediada por otros factores como la experiencia o la motivación.

El papel funcional de las activaciones descritas anteriormente asociadas al procesamiento de la raza no es concluyente. Por un lado podría estar relacionado con diferencias basadas en aspectos físicos o perceptivos de bajo nivel, si bien podría tratarse también de la percepción de amenaza social o peligro que supone el encuentro con el exogrupo. Una forma de tratar de aclarar este asunto sería tratar de introducir estudios con medidas correlacionales de prejuicio. Abordaremos algunos de estos estudios a continuación.

3.2. El estudio de la Raza y el prejuicio desde la cognición social

El interés por el estudio de la raza desde un punto de vista de la psicología social, viene unido a las actitudes y estereotipos que la gente sostiene hacia las personas de otras razas en las interacciones intergrupales y que, en muchas ocasiones llevan al desarrollo de prejuicios hacia los miembros del exogrupo. El término *estereotipo* se puede entender como un conjunto de creencias compartidas socialmente sobre rasgos que son característicos de los miembros de una categoría social, que ayudan a simplificar y sistematizar la información. Según Allport, se trata de una creencia exagerada, favorable o desfavorable, asociada a una categoría (Allport, 1954). El *prejuicio* consiste en formar una opinión en ausencia de experiencia directa con una persona o cosa. Es una actitud (de antipatía en su mayoría, aunque también existe el prejuicio positivo; Dienstbier, 1970) basada en generalizaciones hacia un grupo o un individuo. El prejuicio sería un tipo de actitud en la que se evalúa a un grupo o a sus miembros, en este caso en base a su raza (racismo). La hipótesis del contacto propuso que esto cambia con la experiencia con otras razas (Allport, 1954; Amir, 1969), haciendo referencia a que el contacto que se da en las relaciones intergrupales reduce en alguna medida el prejuicio hacia el grupo en cuestión

Desde la psicología social se han desarrollado diferentes maneras de medir el prejuicio. Aparte de las medidas directas clásicas, que consisten en medidas de auto-informe (e.g. *the Modern Racism Scale*; McConahay, 1986), desde los años 40, una miscelánea de medidas indirectas ponen de manifiesto lo que Greenwald y Banaji (1995) denominaron posteriormente *racismo implícito.* Por un lado, se llevaron a cabo estudios con medidas fisiológicas que comparan respuesta galvánica o tasa cardíaca ante miembros de una raza vs. otra (Rankin & Campbell, 1955; Westie & DeFleur, 1959; Vidulich & Krevanick, 1966). Por otro, a raíz del surgimiento del concepto de cognición implícita (véase Greenwald & Banaji, 1995), una serie de estudios en actitudes y estereotipos ponen de manifiesto procesos automáticos en el prejuicio. Entre ellos encontramos los estudios pioneros de Gaertner y McLaughlin (1983), Dovidio et al., (1986) y Devine (1989), aludiendo al papel de procesos automáticos y no conscientes en estereotipos y prejuicio. Son automáticos en el sentido de que se dan con independencia de actitudes y creencias explícitas. Por ejemplo, es importante el uso de tareas indirectas de decisión léxica y de *priming* en las que se responde más rápidamente a asociaciones estereotípicamente congruentes con la raza en cuestión (Gaertner & McLaughlin, 1983). Entre las medidas indirectas desarrolladas en las últimas tres décadas más usadas en cognición social para medir prejuicio implícito están el test de asociación implícita (IAT; del inglés *Implicit Association Test*) o las tareas de *priming* (Payne, 2001). Aunque, como Stanley y cols. (2008) han señalado, la distinción entre actitudes implícitas y explícitas surgió en las últimas décadas y, en la práctica esta acaba siendo una cuestión metodológica más que teórica (Stanley, Phelps, & Banaji, 2008).

Usando estas medidas, se han encontrado correlaciones entre diferentes correlatos neurales y medidas implícitas de asociación racial. Los primeros estudios de neuroimagen usando presentaciones de caras de gente blanca y negra encontraron activación diferencial en la amígdala, con menor activación ante miembros del endogrupo (Hart, Whalen, Shin, McInerney, Fischer, & Rauch, 2000), que además correlacionaba con medidas indirectas de actitudes raciales (Phelps et al., 2000). En un estudio posterior que extiende estos resultados (Cunningham, Johnson, Raye, Gatenby, Gore, & Banaji, 2004) se comparó el procesamiento automático vs. controlado, encontrando este patrón de activación en la amígdala ante la presentación durante 30 ms de caras de raza negra, que correlacionaba con el sesgo del participante medido mediante el IAT, de manera que una mayor puntuación en IAT para negros predecía una mayor activación en la amígdala para los mismos. Además, un mayor sesgo racial correlacionó de forma significativa con mayor activación en el giro fusiforme (área relacionada con pericia perceptiva) ante las caras de la raza propia comparada con caras de la otra raza. En contraste, presentaciones de 525 ms de caras de raza negra dieron lugar a mayor activación en zonas de la corteza prefrontal y cíngulo anterior que también correlacionó con las puntuaciones en el IAT, (patrón consistente con el estudio de Richeson, Baird, Gordon, Heatherton, Wyland, Trawalter, & Shelton, 2003). La activación en la amígdala la atribuyen a las respuestas emocionales automáticas, que parecen verse potenciadas ante mirada desviada (Richeson et al., 2008), y desaparecen cuando la gente tiene tiempo de procesar el estímulo y ejercer control sobre las actitudes automáticas, controlando así las respuestas de prejuicio ante caras de negros.

En otro estudio, que investigaba los correlatos neurales de diferentes razas con registro electroencefalográfico (He, Johnson, Dovidio, & McCarthy, 2009) se encontraron correlaciones con el IAT. Alrededor de los 170 ms, las diferencias entre caras de la misma raza y de raza negra correlacionaron con el sesgo en asociaciones implícitas. Aunque los datos de estos dos experimentos no son directamente comparables, la conclusión que tratan de extraer apunta a un procesamiento temprano más automático (reflejado en la activación y correlaciones con amígdala, y en la correlación entre la amplitud de ERPs para negros y el IAT), y otro más tardío y controlado, mediado por corteza prefrontal y cingulada. Parece pues que la raza a nivel cerebral se procesa de forma rápida y automática, sin que ello impida poder ejercer un control en una fase más tardía sobre las propias actitudes y evaluaciones.

En el último apartado de esta introducción retomaremos esta revisión mencionando algunos estudios que introducen los sesgos raciales y las medidas implícitas en el contexto de la toma de decisiones y la confianza.

4. *Aspectos variantes en el rostro. La expresión emocional en las relaciones interpersonales.*

"The events I am describing in a brainless creature (the unbrained Paramecium) already contain the essence of the process of emotion that we humans have"

(Damasio, looking for Spinoza)

Hay características identificables del rostro en las cuales nos basamos para guiar nuestras interacciones actuales y futuras con la gente. Entre los aspectos variantes de una cara, se encuentran la dirección de la mirada, los movimientos del habla, y la expresión emocional. Probablemente se trate del estímulo que más varía en el transcurso de un intervalo de tiempo relativamente breve. Piénsese solamente en los movimientos que pueden realizar los ojos en algunos segundos, y que pueden estar comunicando desde un punto de interés al cual es adecuado dirigir la atención en un momento dado, hasta una emoción de sorpresa o incluso el estado de alerta, somnolencia o aburrimiento de la persona en cuestión. Este tipo de información es además de la que más significados se pueden extraer a través no sólo de claves visuales sino también auditivas, a través de las modulaciones del habla. Existe una vasta evidencia de la influencia que esta última tiene en la toma de decisiones y las relaciones interpersonales (Olsson & Ochsner, 2008). Usamos las emociones que percibimos en los demás para predecir qué pasa en su mente y cómo van a actuar, ya que como sugiere una visión evolutiva hemos aprendido a asociar diferentes emociones con significados específicos (Darwin, 1872).

A continuación revisaremos literatura relevante en el ámbito de la toma de decisiones pero antes vamos a hacer una revisión teórica más general de la función de la expresión emocional.

4.1. Funciones de las expresiones emocionales en la interacción social.

Dejando aparte aproximaciones filosóficas, Darwin (1872), fue uno de los pioneros en considerar que las expresiones emocionales eran innatas, universales por tanto entre las diferentes culturas y su función primordial era la de comunicar estados internos. Un siglo después, Ekman y Friesen (1971) abogando también por la universalidad, investigaron si culturas que no han tenido contacto con la alfabetización interpretan el comportamiento facial de la misma manera. Proponen que hay un elemento universal que se manifiesta en unas 6 emociones básicas (rabia, asco, miedo, alegrías, tristeza, sorpresa) reconocibles en todas las culturas. Pero estudios posteriores cuestionaron este supuesto al encontrar casos de culturas aisladas donde no se daba tal universalidad en el reconocimiento de estas emociones (véase Russell, 1994). Así, Podríamos enmarcar las diferentes perspectivas que han tratado de estudiar las expresiones emocionales y su significado a lo largo de un continuo que iría desde adaptaciones evolutivas, innatas y universales, (Darwin 1872; Ekman & Friesen, 1971; Cosmides & Tooby, 2000) hasta constructos aprendidos socialmente, por tanto variantes y supeditadas a la cultura (Fridlund, 1994; Parkinson, 1996; Parkinson et al., 2004).

Una serie de teorías clásicas apuntan a la importancia de la autopercepción de las reacciones corporales desencadenadas por un hecho (James, 1884); según estas, la propiocepción de los cambios fisiológicos y viscerales y su interpretación es lo que provoca a la emoción. La emoción sería producto de una función conjunta del *arousal* autonómico y las atribuciones cognitivas para ese arousal (Schater, 1964), aunque cada autor le da un peso y una secuencia distinta. Así, también podríamos considerar un continuo que iría desde la percepción del cambio fisiológico y el arousal, a los procesos más cognitivos, donde teorías como la del *embodiment*, proponen que al observar una expresión emocional se dispara una respuesta de simulación en nuestros sistemas cerebrales somatosensoriales, motores y de reforzamiento. De manera que la percepción de una expresión emocional se acompaña por los estados corporales y neurales asociados y su correspondiente emoción (Niedenthal, Mermillod, Maringer, & Hess, 2010). Y en la actualidad, la perspectiva del construccionismo propone las emociones como estados emergentes de los diferentes niveles de procesos cerebrales (Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012)

Las emociones son parte de los mecanismos básicos de regulación (Damasio, 2003). En su aspecto básico, son funciones biológicas del sistema nervioso (LeDoux, 1996). Pero, ¿cuál es su maquinaria? Según Damasio (2003), las emociones se construyen a partir de reacciones simples que promueven la supervivencia de un organismo, por lo que el razonamiento y la deliberación no es un requisito del cual dependan. Todos los organismos nacemos con estrategias para resolver *automáticamente* los problemas básicos de la vida (mediante procesos homeostáticos), desde procesos metabólicos hasta gestión de emociones y sentimientos, que a su vez disparan respuestas, desde las más simples a las más complejas de aproximación-rechazo, activación-calma, cooperación-competición, siendo esta última categoría la más relevante para nuestra investigación.

Ha habido diferentes maneras de clasificar las emociones, (véase por ejemplo Ekman, 1992 o Woodworth, 1938). Por su parte Damasio distingue tres tipos de emociones:

El *background* emocional, sería lo equivalente al estado de ánimo que consultamos cuando nos preguntan cómo estás; las emociones primarias o básicas, equivalentes a las 6 emociones primarias de Eckman, que suelen incluirse en esta categoría en base a la consistencia encontrada entre culturas; y emociones sociales, que incluyen vergüenza, culpa, orgullo, admiración, indignación, y provendrían de un conglomerado de reacciones regulatorias básicas y emociones primarias. Parece que la disposición a exhibir una emoción social está bien arraigada en el cerebro, incluso de animales no humanos, listas para implementarse cuando la situación la dispara. Esta última diferenciación está relacionada con el aspecto social de las emociones, fundamental en el estudio de la emoción.

Entre los autores que destacan este aspecto, se encuentra Brian Parkinson (1996; pero véase también Fridlund, 1994; Fischer & Manstead, 2008; Keltner & Kring, 1998; Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012). Parkinson sitúa las causas, consecuencias y funciones de las emociones en un plano social. Según este autor la psicología social es el mejor punto de enfoque para muchos aspectos del estudio de la emoción y para una teoría de la emoción completa (frente a una perspectiva más fisiológica o cognitiva/individual), dando prioridad conceptual a lo social, frente a los procesos cognitivos o las respuestas fisiológicas de la emoción. Enfatiza así los factores sociales, cómo las emociones se producen siempre en un contexto particular que les da su significado y moldean su desarrollo. Revisaremos a continuación algunos de sus planteamientos.

- Una de las principales causas de la emoción son los demás. La gente es la que nos provoca emociones; a su vez, las emociones de los otros tienen consecuencias emocionales en nosotros. Los eventos que causan estados emocionales a menudo alcanzan un significado personal en el transcurso de los encuentros sociales y de las relaciones entre la gente. Por lo general, lo que los otros hacen y dicen es lo que más nos afecta. Tanto más cuando tenemos algún tipo de relación o implicación con ellos; además, la naturaleza de esta conexión interpersonal determina el significado particular del estímulo emocional (las implicaciones afectivas no son las mismas ante un cumplido que venga de tu jefe, de tu empleado, o de alguien que te gusta).

- Las emociones son parte de un proceso social dinámico, ya que median transacciones entre la gente. La presencia de los otros puede tener un efecto facilitador o inhibidor en la expresión emocional. Las emociones de los demás pueden contribuir así a causar emoción, por contagio emocional o por reciprocidad emocional, sin que intervenga una evaluación racional de la conducta de la otra persona. También, el tomar en consideración las reacciones de los demás hacia un objeto o evento es una manera de juzgar la relevancia potencial personal del mismo. Y en la otra dirección, los otros pueden verse afectados por nuestras evaluaciones. Las emociones son maneras de alinear y re-alinear las relaciones interpersonales e intergrupales.

- La expresión facial tiene una función de comunicación, como un medio de transmitir un mensaje a los demás. Parkinson, Fischer y Manstead (2004) proponen que las caras, sobre todo las que expresan significado emocional, se entienden mejor como parte de un contexto interpersonal y no como expresiones individuales, dado que una función clave de muchas expresiones emocionales es la de provocar efectos interpersonales; son procesos orientados a otros. Es decir, parte del propósito de tener una emoción puede ser producir una reacción en alguien. Conforme se desarrolla la interacción, la comunicación emocional va alterando y ajustando *online* el tono de dicha interrelación; de esta manera, la respuesta de los otros redirige las interacciones emocionales en curso. Todo esto en un marco de referencia compartido, tanto a nivel ontogenético, como cultural. Por ejemplo, la respuesta de una persona al enfado de otra depende de la manera en la que se ha socializado, su contexto socio-cultural (la sociedad y la cultura nos enseña cómo interpretar lo que pasa y qué hacer al respecto; como un "sistema de significados") y el tipo de relaciones emocionales tempranas que estableció con sus cuidadores (tipo de apego, etc.).

Pero no en todas las culturas, los contextos sociales o en todas las interacciones, comunica lo mismo una emoción dada. A. J. Fridlund (1994) pone el énfasis en la importancia del contexto social y cultural de la emoción, con una perspectiva más ecológica. Este autor no trata las expresiones emocionales como estados emocionales discretos o fundamentales, y cuestiona que el rostro las exprese de una manera comprendida universalmente, igual en todas las culturas. Se avala para ello en estudios con sociedades aisladas e iletradas (Ekman et al., 1969). Contrasta la psicología clásica de las emociones con su visión "ecológica" de la conducta, donde las expresiones faciales son herramientas sociales para la negociación en los encuentros sociales, son específicas al contexto y al objetivo y surgen de la interacción social, por lo que deben ser interpretadas dentro del contexto. Fridlund argumenta que la relación entre la exhibición de una emoción en el rostro y la emoción que se tiene en un momento dado es adventicia. Por ejemplo, aunque la sonrisa tiene que ver con el hecho de que nos sintamos realmente felices en un momento dado, se puede emitir con significados muy diferentes por razones de cortesía, cariño o diversión, con carácter afiliativo, de dominancia (Niedenthal et al., 2010). Las conductas de exhibición facial "tienen significado en el contexto social donde han sido emitidas […] y reflejan la motivación de cada uno, dentro de un contexto específico de interacción" (Fridlund, 1994, pg.330; véase también Russell, 1994).

La atribución automática de emociones está ligada a la teoría de la mente. Las emociones de los demás están causalmente interrelacionadas con las atribuciones de sus estados mentales; de otra manera, comprender un rostro de alegría sería una tarea ardua (está contento porque consiguió su objetivo, por ejemplo; Saxe et al., 2004).

4.2. Bases neurales en el procesamiento de emociones faciales

Desde la neurociencia social se ha estudiado el procesamiento de las emociones faciales a partir de los sustratos neurales e índices electrofisiológicos subyacentes. Tradicionalmente, el sistema límbico incluyendo la amígdala y la corteza orbitofrontal, abarca en términos generales los circuitos básicos de procesamiento de estímulos emocionales, o el "cerebro emocional" (Papez, 1937; MacLean, 1949).

Estudios con neuroimagen y a nivel electrofisiológico tratan de dilucidar las redes cerebrales encargadas del procesamiento de las emociones, en concreto a la hora de percibirlas en las expresiones faciales. En la década de los 90, algunos estudios apuntaron al procesamiento de emociones específicas en sistemas corticales discretos, de manera que a cada emoción subyacería una *circuitería* o áreas neurales específicas y diferentes, y hay estudios que tienden a precisar qué áreas intervienen o modulan emociones específicas. La principal evidencia viene del campo de la neuropsicología. Por ejemplo, los estudios clásicos muestran que el daño en la amígdala afecta al reconocimiento del miedo (Adolphs, Tranel, Damasio, & Damasio, 1994; Calder, Young, Rowland, Perrett, Hodges, & Etcoff, 1996), aunque a veces daña al reconocimiento conjunto de otras emociones (Rapcsak, Galper, Comer, Reminger, Nielsen, et al., 2000). Esta idea, un tanto localizacionista, en la actualidad está bastante obsoleta. Además, estos datos se obtienen con estudios de caso, que no permiten extraer conclusiones fehacientes, dadas las diferencias individuales, la extensión de la lesión, el tipo de tarea, y la dificultad de hacer generalizaciones a partir de un caso único de funcionamiento anómalo.

En estudios de neuroimagen sin daño adquirido, la evidencia localizacionista más reiterada se ha dado para el procesamiento de las emociones de miedo y asco; se ha enfatizado el rol de la amígdala ante estímulos que producen miedo o son amenazantes (Morris, Frith, Perrett, Rowland, Young et al., 1996; Vuilleumier, Armony, Driver, &

Dolan, 2001), incluso cuando se perciben caras de miedo de manera inconsciente (Morris, Ohman, & Dolan, 1998; Whalen, Rauch, Etcoff, McInerney, Lee, & Jenike, 1998). La ínsula, por su parte, es una zona que está relacionada entre otras cosas con la percepción del propio cuerpo, con la información gustativa, y media en la emoción de asco (Phillips, Young, Senior, Brammer, Andrew, Calder et al., 1997); aunque activación en la ínsula anterior se relaciona con una variedad de tareas que implican conciencia de estados corporales, sin que se de asco per se (Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012). También se ha asociado la corteza orbitofrontal (OFC) a enfado.

En un estudio de Winston y cols. (2003) se obtiene activación en la amígdala ante emociones de asco, miedo, alegría o tristeza, ante expresiones emocionales de alta intensidad, lo que sugiere cierta implicación de esta estructura para estímulos emocionales intensos, independiente de la valencia de los mismos (Winston, O'Doherty, & Dolan, 2003; pero véase Morris, Frith, Perrett, Rowland, Young, Calder, & Dolan, 1996; Blair, Morris, Frith, Perrett, & Dolan, 1999). Si consideramos, por un lado, que la amígdala está implicada en aprendizaje afectivo positivo y procesos de reforzamiento (Baxter & Murria, 2002), y, por otro lado, que la información puede tener acceso a la amígdala a través de dos rutas, una lenta cortical y otra rápida subcortical a través de los colículos (LeDoux, 1996), podría darse un procesamiento a diferentes niveles según el tipo de emoción y de situación estimular. En esta dirección apunta actualmente la neurociencia de las emociones, con una visión más construccionista, dejando de lado la idea que las categorías emocionales discretas pueden ser consistentemente localizadas en distintas regiones cerebrales, y por el contrario, acercándose a la idea de que las emociones son estados mentales que emergen de operaciones psicológicas básicas no específicamente emocionales (Lindquist et al., 2012). En concreto, Lindquist y cols. proponen que los humanos extraemos el significado de las sensaciones corporales que producen los estímulos motivacionalmente salientes (*core affect*) usando las representaciones almacenadas de experiencias pasadas (recuerdos, conocimiento) para extraer el significado (conceptualizando) de la sensación del momento. Así pues, una gama de regiones cerebrales asociadas a la operación de conceptualización entrarían en juego, incluyendo la corteza prefrontal dorsomedial (CPFDM) y ventromedial (CPFVM), el lóbulo temporal medial (LTM), y la corteza cingulada posterior (CCP).

Para estudiar no sólo el dónde sino cuándo comienza a procesarse una emoción, y también para ver el momento en el que el cerebro empieza a detectar las diferencias entre emociones, si las hubiera, una vasta cantidad de estudios de ERP en las últimas décadas han investigado el curso temporal del procesamiento de las expresiones faciales, mediante la presentación de fotografías de caras que muestran diferentes emociones, o expresión neutra (Krolak-Salmon, Fischer, Vighetto, & Mauguiere, 2001); Eimer, Holmes, & McGlone, 2003; Eimer & Holmes, 2007; Batty y Taylor, 2003; Schupp, Öhman, Junghöfer, Weike, Stockburger, & Hamm, 2004; Ashley, Vuilleumier, & Swick, 2004; Utama, Takemoto, Koike, & Nakamura, 2009). Algunos estudios con ERP concluyen que las expresiones faciales emocionales comienzan a modular componentes tempranos, alrededor de los 100 ms, con una positividad frontocentral similar para todas, frente a las expresiones neutras, sugiriendo que no hay estructuras neurales especializadas para la detección de emociones específicas (Eimer, Holmes, & McGlone, 2003; Eimer & Holmes, 2007). Esto concuerda con algunos estudios de enmascaramiento de rostros, los cuales muestran una gran precisión en el procesamiento de la expresión emocional, de forma que exposiciones de 80 ms permiten extraer información suficiente de una cara como para identificar su expresión, aunque no otros aspectos como el género (Aguado, Serrano-Pedraza, & García-Gutiérrez, en prensa).

No obstante, entre los estudios con electroencefalografía existe también alguna evidencia de potenciales evocados específicos asociados a emociones de miedo y asco alrededor de los 200 ms (Ashley, Vuilleumier, & Swick, 2004) y una facilitación en el procesamiento perceptivo de caras amenazantes (Schupp et al., 2004) que comenzaría alrededor de los 100 ms después de la presentación del estímulo. También Batty y Taylor (2003) muestran datos que apoyan redes neurales diferentes para el procesamiento de emociones negativas; mostraron diferentes patrones de activación entre caras emocionales y neutras a partir de los 90 ms en el P1, y que diferenciaban emociones negativas a partir de los 170 ms. Esto apoyaría también una teoría de procesamiento cerebral de expresiones faciales en las que diferentes expresiones emocionales reclutan redes parcialmente disociables.

Hay por tanto una variedad de resultados dependiendo de la metodología y la tarea utilizada, y hay que tener en cuenta variantes como que la presentación se haga por bloques o de manera aleatoria, que la tarea requiera un procesamiento de la emoción directo o incidental, enmascarado o consciente, el papel de la atención al procesar la emoción, etc. Por ejemplo, hay estudios en los cuales desaparecen los efectos de la emoción cuando la atención se dirige a una tarea de discriminación perceptiva (Eimer, Holmes, & McGlone, 2003) o a una tarea de discriminación de género (Krolak-Salmon, et al., 2001), así como otros obtienen un efecto en el curso temporal del procesamiento de caras emocionales en tareas implícitas (Batty & Taylor, 2003).

En lo que sí parecen confluir y concluir toda la investigación en este ámbito es en el rápido y automático procesamiento de las expresiones emocionales (Aguado y otros, en prensa).

5. *La Raza y las emociones como reguladores de la interacción social.*

> “…reactions that lead to racial and cultural prejudices are based in part on the automatic deployment of social emotions evolutionarily meant to detect *difference* in others because difference may signal risk or danger, and promote withdrawal or aggression. [...] We can be wise to the fact that our brain still carries the machinery to react in the way it did in a very different context ages ago"
>
> *Damasio, looking for Spinoza*

Es importante plantearse la función que tiene este rápido procesamiento de la información que proporciona un rostro, y cómo un factor social, invariante de la persona como es la raza, y un factor variante del rostro como es la emoción expresada, modulan la decisión de confiar en nuestra interacción social con los otros. Esto puede estar afectando el comportamiento a muchos niveles sin ser conscientes de ello. Así pues, en esta tesis doctoral vamos a abordar la influencia de la expresión emocional y la raza, desde la perspectiva de la cognición social, en las interacciones sociales a través de un paradigma de teoría del juego como es el Juego de Confianza, el cual describiremos en el apartado siguiente. Nos preguntamos cómo se reflejan las interacciones con personas de otra raza y ante diferentes emociones faciales en los índices electrofisiológicos de procesamiento de caras, y también a nivel comportamental, mediante un paradigma de toma de decisiones.

Una manera en la que estos factores interactúan en la vida cotidiana se da en las relaciones entre grupos sociales, en las que a menudo aparecen emociones negativas. Muchas veces los sentimientos hacia otros dependen principalmente del hecho de pertenecer a un grupo frente a otro (Parkinson et al., 2004), como ocurre con los prejuicios hacia miembros de otra raza. Ansiedad o ira son a menudo emociones experimentadas en la interacción con un exogrupo (véase Dijker, 1987, teoría de la imagen; véase también Smith, 1993, teoría de la autocategorización). Hay una relación entre emociones y actitudes intergrupales, de manera que a veces los individuos experimentan emociones por cosas que han hecho o les han hecho como miembros de un grupo social. En la medida en que uno se clasifica a sí mismo como perteneciente a un grupo, experimentará emociones simplemente por esta clasificación. De cómo modulan estas variables la toma de decisiones trataremos en los subapartados siguientes, así como de los paradigmas dentro de la teoría del juego que se han utilizado para estudiarlo.

5.1. Toma de decisiones: Confianza y Juegos económicos

¿Voy a esa fiesta? ¿Confío en este vendedor? ¿Ayudo a un amigo a pesar del coste que supone para mí? ¿Emprendo un negocio con esta persona? ¿Tengo un hijo con esta pareja? Para responder a estas preguntas, debemos considerar y procesar las alternativas que barajamos y determinar el curso de acción que nos resulte más óptimo, esto es, tomar una decisión. El estudio de la toma de decisiones trata de entender esta habilidad humana. La mayoría de nuestras decisiones, desde las más triviales a las más importantes se hacen en un contexto social, de interacción con los demás, donde no sólo somos dependientes de nuestras propias decisiones y sus consecuencias, sino también de las decisiones concomitantes de los demás y lo que ellas pueden ocasionar a su vez (Rilling & Sanfey, 2011). Para la investigación en toma de decisiones se ha utilizado en las últimas décadas tareas provenientes de la Teoría del Juego, del ámbito de la neuroeconomía.

La teoría del juego se desarrolló a mediados del siglo XX, con autores como Neumann y Morgenstern (1944) y Nash (1950); define la toma de decisiones en una situación social de conflicto (Güth, 1998), en la que se puede actuar de una forma cooperativa o competitiva. Los paradigmas económicos son una manera sencilla de estudiar interacciones permitiendo el control y la manipulación de variables de interés. Algunos de los principales paradigmas son el *Juego del Ultimatum* (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003; Camerer & Thaler, 1995), el *Dilema del Prisionero* (Rilling, Gutman, Zeh, Pagnoni, Berns, & Kilts, 2002) y el *Juego de Confianza* (Berg, Dickhaut & McCabe, (1995).

Para el desarrollo de esta tesis, a lo largo de la serie experimental hemos utilizado una adaptación del Juego de Confianza, por resultar adecuado para estudiar reciprocidad y toma de decisión acerca de confiar o no en los demás. En este paradigma hay dos jugadores, el participante y su pareja de juego. Los participantes ven caras de diferentes personas en el ordenador, que son sus hipotéticos compañeros de juego. Su tarea es decidir cooperar o no con ellos, compartiendo o no una cantidad del dinero en juego. En cada ensayo, el participante comienza con una cantidad de dinero y tiene que decidir si lo comparte o no con su pareja de juego. Si se queda con el dinero (no coopera) la pareja de juego se queda sin nada; si por el contrario decide cooperar y comparte su dinero, el otro jugador recibe la cantidad compartida multiplicada por *x*, luego supuestamente esta pareja decide si ser recíproco en cuyo caso la suma se divide entre los dos, o si se queda con todo, con lo cual el participante lo pierde todo.

En este tipo de tareas, se observa cómo la toma de decisiones parece ser menos estratégica y racional de lo que originariamente se propuso desde la teoría del juego (Neumann & Morgenstern, 1944; Nash, 1950). El comportamiento se desvía de lo predicho por estos modelos; en el caso del Juego de Confianza, hay muchos factores relacionados con la pareja de juego que influyen en la confianza (véase Tzieropoulos, 2013, para una revisión reciente). Por ejemplo, en un estudio con este paradigma, los participantes cooperaban más con las caras más atractivas, (Chen, Zhong, Zhang, Li, Zhang, Tan, & Li, 2012). Así, la gente se basa en factores de tipo personal para decidir, donde la confianza es muy relevante.

Los juicios sobre confianza en los demás se pueden hacer de dos formas no excluyentes. En un primer momento cuando aún no tenemos información de la persona, ocurren de manera muy rápida y automática, a través de la cara de los demás y la primera impresión que de ellos extraemos, haciendo juicios generales de confianza (Van‘t Wout & Sanfey, 2008; Todorov, Said, Engell, & Oosterhof, 2008). Por ejemplo, la gente puede basar sus primeras impresiones en presentaciones de caras a los 39 s, (Bar et al., 2006) que parecen activar esquemas de evaluación de manera rápida y automática. Posteriormente, tras haber interactuado con ellos, obtenemos evidencia acerca de confiar o no en ellos por su comportamiento o incluso sabiendo de su reputación a través de otros (Frith & Singer, 2008).

En las últimas décadas, se ha comenzado a investigar correlatos neurales de la toma de decisiones usando tareas que provienen de la Teoría del Juego. La investigación más actual desde la perspectiva de la neurociencia en toma de decisiones proviene de estudios con IRMf (Rilling & Sanfey, 2011). Se han puesto de manifiesto una serie de redes cerebrales implicadas en la toma de decisiones y muchas son coincidentes con la circuitería implicada en ToM, que mencionamos al principio de esta introducción. De hecho, estas tareas sociales de toma de decisiones tienen una estrecha relación con cómo procesamos las intenciones y acciones de los demás, estrechamente ligado al hecho de hacer inferencias y predicciones para decidir si confiar o no en alguien.

Una estructura fundamental en juicios de confianza es la amígdala. Daños bilaterales afectan la capacidad de juzgar las caras de gente que no es de confianza (Adolphs, Tranel & Damasio, 1998). Asimismo, en poblaciones normales las caras que no son de confianza activan la amígdala (Winston, Strange, O'Doherty, & Dolan, 2002). Muchos estudios han encontrado que la respuesta de esta estructura se incrementa conforme la desconfianza en la cara de la otra persona aumenta (Engell, Haxby, & Todorov, 2007; Winston et al., 2002). Sin embargo, dado el hallazgo de una mayor activación ante caras alegres que neutras en algunas investigaciones (Winston et al., 2003; Yang, Menon, Eliez, Blasey, White, Reid, Gotlib, & Reiss, 2002) sugiere una respuesta no lineal, con activaciones de la amígdala más intensas ante caras extremas tanto de confianza como de desconfianza (Todorov et al., 2008). Esto iría acorde con la propuesta expuesta anteriormente respecto a la implicación en emoción de la amígdala.

Otro sistema también implicado en toma de decisiones es el sistema mesolímbico dopaminérgico, tradicionalmente asociado a recompensa (Schultz, 1998); concretamente, el estriado ventral juega un papel fundamental, con cambios de activación directamente relacionados con la magnitud de la recompensa, tanto de tipo social como no social (Montague & Berns, 2002). Estudios con el Juego de confianza han puesto de manifiesto que la actividad en el estriado está relacionada con decisiones; se da un incremento en la activación cuando se elige cooperar vs. no cooperar. En un estudio con *hiperscanning*, tras aprender a predecir la reciprocidad de la pareja de juego, la actividad en el caudado se desplazó en el tiempo desde el momento de recibir el resultado al momento de la decisión (King-Casas, Tomlin, Anen, Camerer, Quartz, & Montague, 2005). Otras estructuras como la ínsula anterior, implicada en empatía, muestra mayor activación conforme la oferta que se hace en el *Juego del Ultimátum* es más injusta (Sanfey et al., 2003). La corteza prefrontal medial (CPFm), la corteza cingulada posterior (CCP) y la unión temporoparietal (TPJ), son otras áreas implicadas en las decisiones de confianza así como en ToM (Figura 2; véase Rilling & Sanfey, 2011 para otras zonas implicadas en confianza).

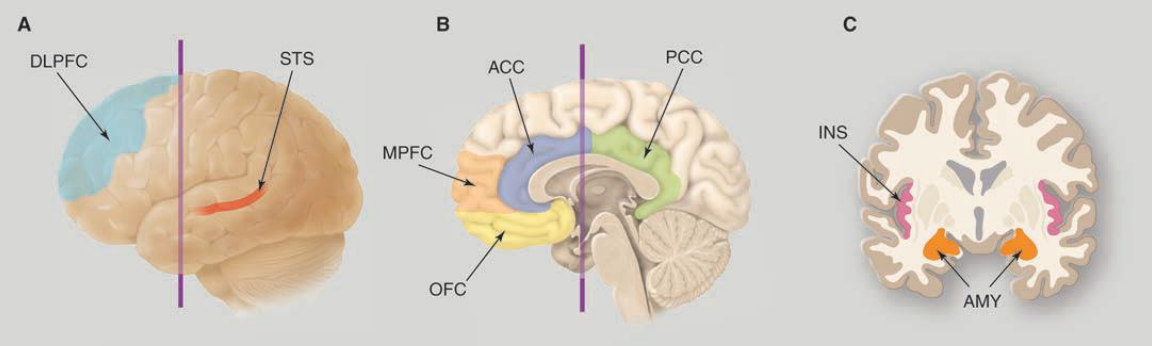


Figura 2. Mapa de las áreas cerebrales que se activan normalmente en estudios de toma de decisiones social (de Sanfey, 2007). Nótese que algunas de estas áreas están también asociadas con procesos de ToM.

También se han realizado registros duales de EEG; en esta línea, un estudio de De Vico Fallani y cols. (2010) registró EEG en parejas de personas mientras jugaban al dilema del prisionero, y mediante el análisis de bandas de frecuencias, mostró la posibilidad de predecir interacciones de no cooperación durante la fase de toma de decisiones; la conducta de no cooperación de una pareja de jugadores estaba asociada a una interacción menor entre las actividades de las áreas corticales de ambos participantes (véase Konvalinka & Roepstorff, 2012, para una revisión). Esta metodología permite observar si a partir de la actividad determinada en un cerebro se predice actividad en otro, y ver 2 cerebros como un único sistema con propiedades emergentes cualitativamente diferentes a la de uno por separado.

5.2. Raza y toma de decisiones

Un factor que modula la toma de decisiones son las diferencias intergrupales y en concreto la raza. En los últimos años, algunos estudios en el marco de la teoría del juego han contemplado la raza para estudiar cómo influye en la toma de decisiones en contextos sociales. La raza es uno de los factores que moldean las decisiones de confianza en los demás, ya que estas decisiones se sustentan en ocasiones sobre las actitudes implícitas y estereotipos que tenemos acerca de los miembros de otras categorías sociales (Stanley, Sokol-Hessner, Banaji, & Phelps, 2011). Además, ante ciertas circunstancias experimentamos emociones no como individuos sino como miembros de un grupo social y hay una tendencia a favorecer al endogrupo frente al exogrupo (Brewer, 1998). Muchas veces los sentimientos hacia otros dependen principalmente del hecho de pertenecer a un grupo social frente a otro (Parkinson et al, 2004); piénsese como claro ejemplo en los seguidores de un equipo de futbol.

En el contexto del juego de confianza hay algunos estudios que muestran efectos comportamentales de raza (Fershtman & Gneezy, 2001), en los que se confía menos con miembros de la otra etnia que de la propia. Algunos autores amplían esta perspectiva y ponen de manifiesto correlaciones entre las actitudes raciales implícitas (medidas mediante el IAT) y las respuestas de cooperación en un juego de confianza (Stanley et al., 2011); en este estudio evaluaban como más “confiables” a los miembros del grupo hacia los cuales las actitudes implícitas eran más favorables. Asimismo, los individuos que mostraban un sesgo pro-blancos en el IAT tendían a ofrecer más dinero a parejas de juego blancas que a negras, y viceversa (Stanley et al., 2011), aunque, en general, no hubo mayores inversiones con miembros de una raza que de otra.

5.3. Emoción y toma de decisiones

La emoción es otro factor importante que modula la toma de decisiones, hay alguna investigación que lo pone de manifiesto (Scharlemann, Eckel, Kacelnik, & Wilson, 2001; Bechara, 2004). Ya que la comunicación de emociones dice mucho sobre las intenciones del que las expresa, los humanos inferimos y creamos expectativas basándonos en las expresiones emocionales de los demás. Muchos estudios versan sobre cómo las emociones afectan en el contexto de la toma de decisiones, aunque la mayoría se han centrado en las emociones sentidas por la persona que toma la decisión y no tanto en la percepción de expresiones emocionales en los demás. Así, es bien sabido que, expresiones de felicidad están ligadas a consecuencias positivas, mientras que expresiones de enfado se asocian a resultados negativos (Darwin, 1872; Cosmides & Tooby, 2000)). También se han estudiado las emociones en el contexto de los juegos económicos; cómo usamos las emociones que los demás expresan para anticipar su comportamiento más probable. En este aspecto, las regiones de procesamiento emocional que están implicadas en juegos de reciprocidad, tienen que ver con el sistema de recompensa, como la corteza orbitofrontal, putamen o el estriado ventral (Marchant & Frith, 2009). Son pioneros los estudios de Antonio Damasio con pacientes con daño en CPFVM. Estos pacientes parecen no aprender de las consecuencias negativas de sus acciones, recurriendo una y otra vez en ellas, a pesar de que sus habilidades intelectuales están preservadas. También se ven alteradas su capacidad de responder a situaciones emocionales, mostrando un afecto plano, lo que llevó a postular que quizá era su incapacidad de hacer uso de las emociones lo que llevaba a los déficits observados en la toma de decisiones en ámbitos personales y financieros de la vida diaria (Naqvi, Shiv, & Bechara, 2006). Esto llevó a postular la hipótesis del marcador somático (Bechara & Damasio, 2005); un punto de partida para entender cómo la capacidad para tomar decisiones se relaciona con la procesos emocionales básicos.

Por lo general, las personas cooperan más a menudo cuando observan una expresión sonriente en otras personas que cuando estas muestran una expresión de enfado, ya que se confía más en gente sonriente que no sonriente (Scharlemann et al, 2001), Hay estudios que muestran que ante una cara de confianza se disparan expectativas y emociones de recompensa (Fehr & Camerer, 2007), lo que sugiere un solapamiento entre áreas cerebrales relacionadas con procesamiento de la recompensa y procesamiento de la emoción. Pero no solamente o no siempre las expresiones de los demás nos proporcionan información fiable sobre cómo se comportarán. El conocimiento previo sobre su fiabilidad, y experiencias pasadas con ellos ayudan a predecir comportamientos futuros (Frith & Frith, 2012). A través de las interacciones actualizamos este conocimiento, que llega a ser más importante que lo que nos comunica su expresión facial (Todorov, Gobbini, Evans, & Haxby 2007).

5.4. Aprendiendo de las consecuencias: La negatividad asociada al error (FRN)

Aprender de las consecuencias de las propias acciones es un mecanismo esencial de los individuos para sobrevivir y adaptarse al medio. Para estudiar los procesos básicos que subyacen a este mecanismo la neurociencia cognitiva estudia qué ocurre en el cerebro cuando una persona procesa información subsecuente a una conducta, esto es, una retroalimentación o *feedback* determinado. El sustrato material a nivel más biológico se remonta a los estudios de respuestas neuronales dopaminérgicas en el sistema medial cerebral (Schultz, 1998, 2002), un sistema implicado en el procesamiento de la recompensa, ante estímulos con significado motivacional. Además, este mecanismo está a la base del aprendizaje clásico e instrumental. Estas neuronas liberan el neurotransmisor dopamina en el estriado y corteza frontal, principalmente, y están activas mientras ocurre el aprendizaje. Y esto se ve reflejado en la manera en que responden, de manera que las respuestas dopaminérgicas dependen de la ocurrencia del evento y de la predicción del mismo: la respuesta es positiva cuando el reforzamiento ocurre sin ser predicho; es nula cuando el refuerzo ocurre tal y como es predicho; y la respuesta es negativa cuando un evento que se predice no ocurre, lo que se ha denominado como un error en la predicción de reforzamiento o *reward prediction error.*

Las predicciones que hacemos sobre lo que todavía no ha ocurrido son por tanto fundamentales al tomar decisiones, pues nos permiten evaluar mentalmente las diferentes posibilidades, integrando y haciendo uso del conocimiento previo que tenemos junto a las posibilidades de ganancias y pérdidas de cada elección.

Quizás uno de los primeros índices cerebrales electrofisiológicos asociados al procesamiento de una consecuencia o resultado derivado de una acción es la negatividad relacionada con el feedback o FRN (del inglés, *feedback-related negativity*). Este potencial se pone de manifiesto cuando realizamos el registro electroencefalográfico (EEG) en paradigmas de toma de decisiones u otro tipo de tareas que impliquen una acción por parte del sujeto y el procesamiento del resultado obtenido. La negatividad relacionada con el feedback (FRN) proporciona un índice electrofisiológico de la valencia de la consecuencia obtenida, con una mayor negatividad para resultados desfavorables que favorables, o que son peor de lo esperado. Esto permite observar diferencias el curso temporal de la actividad eléctrica asociada a la ocurrencia de un evento, en el caso que nos concierne, un feedback. En este aspecto, en esta tesis hemos indagado en las posibles modulaciones de la raza y la expresión emocional sobre la FRN, así como en el papel de las expectativas acerca de la gente de la cual se recibe dicho feedback.

Concluyendo esta sección introductoria, el desarrollo de la neurociencia cognitiva ha hecho que su incursión en las últimas décadas en el ámbito de la cognición social acerque ámbitos que el siglo pasado resultaban difíciles de congregar como son cognición y emoción, los sustratos neurales y el prejuicio o la toma de decisiones y sus modulaciones más allá de lo racional. En el siguiente capítulo expondremos la aproximación experimental y los objetivos de esta tesis.

REFERENCIAS

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

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature,* 393, 470-474.

Aguado, L., Serrano-Pedraza, I., & García-Gutiérrez, A. A comparison of backward masking of faces in expression and gender identification (in press).

Allport, G. (1954). The nature of prejudice. Reading, MA: Perseus Book Publishing.

Amir, Y. (1969). Contact hypothesis in ethnic relations. *Psychological Bulletin*, *71*, 319-342.

Arnold, M.B. (1960). *Emotion and personality*. New York, NY, US: Columbia University Press.

Ashley, V., Vuilleumier, P., & Swick, D. (2004). Time course and specificity of event-related potentials to emotional expressions. *Neuroreport*, *15*, 211-216.

Atkinson, A. P., & Adolphs, R. (2011). The neuropsychology of face perception: beyond simple dissociations and functional selectivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 1726-1738.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*, 269-278.

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: on the automaticity of higher mental processes. *Psychological bulletin*, *126*, 925-945.

Bargh, J. A., & Williams, E. L. (2006). The automaticity of social life. *Current directions in psychological science*, *15*, 1-4.

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT press.

Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research*, *17*, 613-620.

Berry, D. S., & McArthur, L. Z. (1985). Some components and consequences of a babyface. *Journal of personality and social psychology*, *48*, 312.

Bayliss, A. P., & Tipper, S. P. (2005). Gaze and arrow cueing of attention reveals individual differences along the autism spectrum as a function of target context. *British Journal of Psychology*, *96*, 95-114.

Baxter, M. G., & Murray, E. A. (2002). The amygdala and reward. *Nature Reviews Neuroscience*, *3*, 563-573.

Bechara, A. (2004). The role of emotion in decision-making: evidence from neurological patients with orbitofrontal damage. *Brain and cognition*, *55*, 30-40.

Bechara, A., & Damasio, A. (2005). The somatic marker hypothesis: A neural theory of economic decision-making. *Games and Economic Behavior, 52*, 336-372.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience*, *8*, 551-565.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*, 122-142.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS One*, *3*, e3022.

Blair, R. J. R., Morris, J. S., Frith, C. D., Perrett, D. I., & Dolan, R. J. (1999). Dissociable neural responses to facial expressions of sadness and anger. *Brain*, *122*, 883-893.

Brewer, M. B. (1988). A dual process model of impression formation. In R. S. Wyer & T. K. Srull (Eds.), *Advances in social cognition* (Vol. 1, pp. 1-36). Hillsdale, NJ: Erlbaum.

Brewer, M. B., & Brown, R. J. (1998). *Intergroup relations*. McGraw-Hill.

Brigham, J. C., & Barkowitz, P. (1978). Do "They all look alike?" The Effect of Race, Sex, Experience, and Attitudes on the Ability to Recognize Faces1. *Journal of Applied Social Psychology*, *8*, 306-318.

Brothers, L. (1990). The social brain: a project for integrating primate behaviour and neurophysiology in a new domain. *Concepts in Neuroscience, 1*, 27-51.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology*, *77*, 305-327.

Brunet, E., Sarfati, Y., Hardy-Bayle, M. C., Decety, J. (2000): A PET investigation of the attribution of intentions with a nonverbal task. *Neuroimage, 11*, 157–166.

Buck, R. (1984). *The communication of emotion* (pp. 101-5). New York: Guilford Press.

Calarge, Andreasen, & O'Leary, (2003): Visualizing how one brain understands another: A PET study of theory of mind. *American Journal of Psychiatry, 160*, 1954-1964.

Caldara, R., Thut, G., Servoir, P., Michel, C. M., Bovet, P., & Renault, B. (2003). Face versus non-face object perception and the 'other-race'effect: a spatio-temporal event-related potential study. *Clinical Neurophysiology*, *114*, 515-528.

Calder A. J, Young AW, Rowland D, Perrett D. I, Hodges J. R., Etcoff N. L. (1996). Facial emotion recognition after bilateral amygdala damage: differentially severe impairment of fear. *Cogn Neuropsychol, 13*, 699-745.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*, 641-651.

Camerer, C., & Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives*, *9*, 209-219.

Cañadas, E., Rodríguez-Bailón, R., Milliken, B., & Lupiáñez, J. (2013). Social categories as a context for the allocation of attentional control. *Journal of Experimental Psychology: General, 142*, 934-943.

Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human brain mapping*, *30*, 2313-2335.

Chen, J., Zhong, J., Zhang, Y., Li, P., Zhang, A., Tan, Q., & Li, H. (2012). Electrophysiological correlates of processing facial attractiveness and its influence on cooperative behavior. *Neuroscience Letters*, *517*, 65-70.

Cloutier, J., Turk, D. J., & Neil Macrae, C. (2008). Extracting variant and invariant information from faces: The neural substrates of gaze detection and sex categorization. *Social Neuroscience*, *3*, 69-78.

Cohen, L. B., & Cashon, C. H. (2001). Do 7-month-old infants process independent features or facial configurations?. *Infant and child development*, *10*, 83-92.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. *Handbook of emotions*, *2*, 91-115.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science*, *15*, 806-813.

De Renzi, E. (1997). Prosopagnosia. In: Behavioral neurology and neuropsychology (Feinberg TE, Farah MJ, eds), pp 245-255. New York: McGraw-Hill.

De Rivera, J. (1984). The structure of emotional relationships. *Review of Personality & Social Psychology*.

Damasio, A. (2003). Looking for Spinoza: Joy. *Sorrow, and the Feeling Brain*.

Damasio A. R, Damasio H, & Van Hoesen G. W. (1982). Prosopagnosia: anatomic basis and behavioral mechanisms. *Neurology, 32*, 331-41.

Darwin, C. (1872). The Expression of Emotions in Man and Animals. London: John Murray

De Vico Fallani, F., Nicosia, V., Sinatra, R., Astolfi, L., Cincotti, F., Mattia, D., Wilke, C., Doud, A., Latora, V., He, B., & Babiloni, F. (2010). Defecting or not defecting: how to “read” human behavior during cooperative games by EEG measurements. *PLoS One, 5*, e14187.

De Waal, F. B. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annual review of psychology*, 59, 279-300.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56,* 5-18.

Dienstbier, R. A. (1970). Positive and negative prejudice: Interactions of prejudice with race and social desirability1. *Journal of Personality*, *38*(2), 198-215.

Dijker, A. J. (1987). Emotional reactions to ethnic minorities. *European Journal of Social Psychology*, *17*, 305-325.

Dovidio, J. F., Evans, N.E., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology, 22,* 22-37.

Dzhelyova, M. P., Ellison, A., & Atkinson, A. P. (2011). Event-related repetitive TMS reveals distinct, critical roles for right OFA and bilateral posterior STS in judging the sex and trustworthiness of faces. *Journal of cognitive neuroscience*, *23*, 2782-2796.

Eberhardt, J. L. (2005). Imaging race. *American Psychologist*, *60*, 181-190.

Eimer, M. (2000). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clinical neurophysiology*, *111*, 694-705.

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia*, *45*, 15-31.

Eimer, M., Holmes, A., & McGlone, F. P. (2003). The role of spatial attention in the processing of facial expression: an ERP study of rapid brain responses to six basic emotions. *Cognitive, Affective, & Behavioral Neuroscience*, *3*, 97-110.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*, 169-200.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*, 86-88.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, *17*, 124-129.

Engell, A. D, Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: automatic coding of face properties in human amygdala. *Journal of Cognitive Neuroscience, 19*, 1508-19.

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in cognitive sciences*, *11*, 419-427.

Feingold, G. A. (1914). Influence of Environment on Identification of Persons and Things. *J. Am. Inst. Crim. L. & Criminology*, *5*, 39-51.

Fershtman, C., & Gneezy, U. (2001). Discrimination in a segmented society: an experimental approach. *Quarterly Journal of Economics, 116*, 351-377.

Fischer, A. H., & Manstead, A. S. (2008). Social functions of emotion. *Handbook of emotions*, *3*, 456-468.

Fiske, S. T. & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology, 23*, 1-74.

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. Academic Press.

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual review of psychology*, *63*, 287-313.

Frith, C. D., & Singer, T. (2008). The role of social cognition in decision making. *Philos Trans R Soc Lond B Biol Sci. 363*, 3875-3886.

Gaertner, S. L., & McLaughlin, J. E. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly, 46,* 23-30.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in cognitive sciences*, *7*, 77-83.

Gallotti, M., & Frith, C. D. (2013). Social cognition in the we-mode. *Trends in cognitive sciences, 17,* 160-165.

Gauthier, I., & Logothetis, N. K. (2000). Is face recognition not so unique after all?. *Cognitive Neuropsychology*, *17*, 125-142.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature neuroscience*, *3*, 191-197.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform'face area'increases with expertise in recognizing novel objects. *Nature neuroscience*, *2*, 568-573.

George, N., Evans, J., Fiori, N., Davidoff, J., & Renault, B. (1996). Brain events related to normal and moderately scrambled faces. *Cognitive Brain Research*, *4*, 65-76.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior, 24*, 153-72.

Golby, A. J., Gabrieli, J. D. E., Chiao, J. Y., & Eberhardt, J.L. (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience, 4*, 845-850.

Gothard, K. M., Battaglia, F. P., Erickson, C. A., Spitler, K. M., & Amaral, D. G. (2007). Neural responses to facial expression and face identity in the monkey amygdala. *Journal of Neurophysiology*, *97*, 1671-1683.

Gray, J. A. (1971). Sex differences in emotional behaviour in mammals including man: endocrine bases. *Acta Psychologica*, *35*, 29-46.

Greenwald, A.G., & Banaji, M.R. (1995). Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes. *Psychological review, 102*, 4-27.

Güth, W. (1998). Are rational cost expectations evolutionarily stable? *IFO-Studien: Zeitschrift für empirische Wirtschaftsforschung, 44*, 1-13.

Hari, R., Himberg, T., Nummenmaa, L., Hämäläinen, M., & Parkkonen, L. (2013). Synchrony of brains and bodies during implicit interpersonal interaction. *Trends in cognitive sciences*, 17, 105-106.

Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs ingroup face stimuli*. NeuroReport, 11,* 2351-2354.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, *4*, 223-233.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological psychiatry*, *51*, 59-67.

Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron*, *22*, 189-199.

He, Y., Johnson, M. K., Dovidio, J. F., & McCarthy, G. (2009). The relation between race-related implicit associations and scalp-recorded neural activity evoked by faces from different races. *Social Neuroscience*, *4*, 426-442.

Herrmann, M. J., Aranda, D., Ellgring, H., Mueller, T. J., Strik, W. K., Heidrich, A. & Fallgatter, A. J. (2002). Face-specific event-related potential in humans is independent from facial expression. *International Journal of Psychophysiology, 45,* 241-244.

Herrmann, M. J., Ehlis, A. C., Ellgring, H., & Fallgatter, A. J. (2005). Early stages (P100) of face perception in humans as measured with event-related potentials (ERPs). *Journal of neural transmission*, *112*, 1073-1081.

Herrmann, M. J., Schreppel, T., Jäger, D., Koehler, S., Ehlis, A. C., & Fallgatter, A. J. (2007). The other-race effect for face perception: an event-related potential study. *Journal of neural transmission*, *114*, 951-957.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature neuroscience*, *3*, 80-84.

Horstmann, G. (2003). What do facial expressions convey: Feeling states, behavioral intentions, or actions requests?. *Emotion*, *3*, 150-166.

Ishai, A., Schmidt, C., F., & Boesiger, P. (2005). Face perception is mediated by a distributed cortical network. *Brain Research Bulletin, 67*, 87-93.

Ito, T. A., & Bartholow, B. D. (2009). The neural correlates of race. *Trends in cognitive sciences*, *13*, 524-531.

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of personality and social psychology*, *85*, 616-626.

James, W. (1884). II.—What is an emotion?. *Mind*, 34, 188-205.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The journal of Neuroscience, 17*, 4302-4311.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature neuroscience*, *3*, 759-763.

Keating, C. F., Mazur, A., & Segall, M. H. (1977). Facial gestures which influence the perception of status. *Sociometry, 40,* 374-378.

Keltner, D., & Kring, A. M. (1998). Emotion, social function, and psychopathology. *Review of General Psychology*, *2*, 320-342.

Kennedy, D. P., & Adolphs, R. (2012). The social brain in psychiatric and neurological disorders. *Trends in cognitive sciences, 16*, 559-572.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science, 308*, 78-83.

Kluver, H., & Bucy, P. C. (1939). Preliminary analysis of functions of the temporal lobes in monkeys. *Archives of neurology and psychiatry*, *42*, 979-1000.

Konvalinka, I., & Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction?.*Frontiers in human neuroscience*, *6*. 215.

Krolak- Salmon, P., Fischer, C., Vighetto, A., & Mauguiere, F. (2001). Processing of facial emotional expression: spatio- temporal data as assessed by scalp event- related potentials. *European Journal of Neuroscience*, *13*, 987-994.

Kubota, J. T., Banaji, M. R., & Phelps, E. A. (2012). The neuroscience of race. *Nature neuroscience*, *15*, 940-948.

Lazarus, R. S. (1991). Cognition and motivation in emotion. *American psychologist*, *46*, 352-367.

Leslie, A. (1987). Pretense and representation: the origins of “theory of mind.” *Psychological Review, 94*, 412-26.

LeDoux JE (1996): The Emotional Brain. New York: Simon and Schuster.

Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General, 129*, 559-574.

Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annual Review of Psychology,* 58, 259-89.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, *35*, 121-143.

Marchant, J. L., & Frith, C. D. (2009). Social cognition. In L. R. Squire (Ed.), Encyclopedia of Neuroscience (pp. 27-30). Oxford: Elsevier Academic Press.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Science, 6,* 255-260.

MacLean, P. D. (1949). Psychosomatic disease and the" visceral brain" recent developments bearing on the papez theory of emotion. *Psychosomatic Medicine*, *11*, 338-353.

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences*, *98*, 11832-11835.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, *9*, 605-610.

McConahay, J. B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In S. L. Gaertner & J. F. Dovidio (Eds.), *Prejudice, discrimination, and racism: Theory and research* (pp. 91-126). New York: Academic Press.

McNeil, J. E., & Warrington, E. K. (1993). Prosopagnosia: A face-specific disorder. *The Quarterly Journal of Experimental Psychology*, *46*, 1-10.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*, 3-35.

Michel, C., Rossion, B., Han, J., Chung, C. S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science*, *17*, 608-615.

Montague, P. R., & Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron, 36*, 265-84.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., & Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature 383,* 812-815.

Morris, J.S., Ohman, A., Dolan, R.J. (1998). Conscious and unconscious emotional learning in the human amygdala. Nature 393, 467-470.

Naqvi, N., Shiv, B., & Bechara, A. (2006). The role of emotion in decision making a cognitive neuroscience perspective. *Current Directions in Psychological Science*, *15*, 260-264.

Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, *36*, 48-49.

Neumann, J. V., & Morgenstern, O. (1944). Theory of Games and Economic Behavior.

Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences*, *33*, 417- 433.

Ochsner, K. N., &, Lieberman, M. D. (2001). The emergence of social cognitive neuroscience. *American Psychologist,* 56, 717-734.

Olsson, A., Ochsner, K.N. (2008). The role of social cognition in emotion. *Trends in Cognitive Sciences, 12*, 65-71.

Papez, J. W. (1937). A proposed mechanism of emotion. *Archives of neurology and psychiatry*, *38*, 725-743.

Payne, B.K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality Social Psychology, 81*, 181-192.

Parkinson, B. (1996). Emotions are social. *British journal of psychology*, *87*, 663-683.

Parkinson, B., Fischer, A. H., & Manstead, A. S. (2004). *Emotion in social relations: Cultural, group, and interpersonal processes*. Psychology Press.

Perrett, D. I., Mistlin, A. J., & Chitty, A.J. (1987). Visual cells responsive to faces. *Trendy in Neuroscience, 10,* 358-364.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, *12*, 729-738.

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., ... & David, A. S. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, *389*, 495-498.

Polezzi, D., Daum, I., Rubaltelli, E., Lotto, L., Civai, C., Sartori, G., & Rumiati, R. (2008). Mentalizing in economic decision-making. *Behavioural brain research*, *190*, 218-223.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, *1*, 515-526.

Pribram, K. H., & McGuinness, D. (1975). Arousal, activation, and effort in the control of attention. *Psychological review*, *82*, 116-149.

Rapcsak, S.Z., Galper, S.R., Comer, J.F., Reminger, S.L., Nielsen, L., Kaszniak, A.W., Verfaellie, M., Laguna, J.F., Labiner, D.M., Cohen, R.A., 2000. Fear recognition deficits after focal brain damage: a cautionary note. *Neurology 54*, 575-581.

Rankin, R. E., & Campbell, D. T. (1955). Galvanic skin response to Negro and white experimenters. *The Journal of Abnormal and Social Psychology*, *51*, 30-33.

Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S., & Shelton, J. N. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nature neuroscience*, *6*, 1323-1328.

Richeson, J. A., Todd, A. R., Trawalter, S., & Baird, A. A. (2008). Eye-gaze direction modulates race-related amygdala activity. *Group Processes & Intergroup Relations*, *11*, 233-246.

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron*, *35*, 395-405.

Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual review of psychology*, *62*, 23-48.

Rossion, B., Delvenne, J. F., Debatisse, D., Goffaux, V., Bruyer, R., Crommelinck, M., & Guérit, J. M. (1999). Spatio-temporal localization of the face inversion effect: an event-related potentials study. *Biological psychology*, *50*, 173-189.

Rossion, B., Gauthier, I., Tarr, M.J., Despland, P.A., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. *NeuroReport*, *11*, 69-74.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological bulletin*, *115*, 102-141.

Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature neuroscience*, *7*, 499-500.

Sanfey, A.G. (2007). Social decision-making: insights from game theory and neuroscience. *Science,* 318, 598-602.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game.*Science*, *300*, 1755-1758.

Saxe, R. (2006). Uniquely human social cognition. *Current opinion in neurobiology*, *16*, 235-239.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, *55*, 87-124.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in “theory of mind”. *Neuroimage*, *19*, 1835-1842.

Saxe, R., & Powell, L. J. (2006). It's the Thought That Counts Specific Brain Regions for One Component of Theory of Mind. *Psychological Science*, *17*, 692-699.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, *43*, 1391-1399.

Schachter, S. (1964). The interaction of cognitive and physiological determinants of emotional state. *Advances in experimental social psychology*, *1*, 49-80.

Scharlemann, J. P., Eckel, C. C., Kacelnik, A., & Wilson, R. K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology*, *22*, 617-640.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of neurophysiology*, *80*, 1-27.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, *36*, 241-263.

Schupp, H. T., Öhman, A., Junghöfer, M., Weike, A. I., Stockburger, J., & Hamm, A. O. (2004). The facilitated processing of threatening faces: an ERP analysis. *Emotion*, *4*, 189.

Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *The Neuroscientist*, *17*, 18-24.

Singer, T., Seymour, B., O’Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science 303*, 1157-1162.

Singer, T., Seymour, B., O’Doherty, J., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439*, 466-469.

Smith, E. R. (1993). Social identity and social emotions: Toward new conceptualizations of prejudice. In D. M. Mackie & D. L. Hamilton (Eds.) Affect, cognition, and stereotyping: Interactive processes in group perception. San Diego, CA: Academic Press.

Stahl, J., Wiese, H., & Schweinberger, S. R. (2008). Expertise and own-race bias in face processing: an event-related potential study. *Neuroreport*, *19*, 583-587.

Stanley, D., Phelps, E., & Banaji, M. (2008). The neural basis of implicit attitudes. *Current Directions in Psychological Science*, *17*, 164-170.

Stanley, D. A., Sokol-Hessner, P., Banaji, M. R., & Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *PNAS, 108*, 7710-7715.

Stanley, D. A., Sokol-Hessner, P., Fareri, D. S., Perino, M. T., Delgado, M. R., Banaji, M. R., & Phelps, E. A. (2012). Race and reputation: perceived racial group trustworthiness influences the neural correlates of trust decisions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 744-753.

Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. *Advances in experimental social psychology*, *11*, 249-288.

Todorov, A., Gobbini, M. I., Evans, K. K., & Haxby, J. V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia, 45*, 163-173.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. Trends in cognitive sciences, 12, 455-460.

Tskhay, K. O. & Rule, N. O. (2013). Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Personality and Social Psychology Review, 17*, 72-86.

Tzieropoulos, H. (2013). The Trust Game in neuroscience: A short review. *Social neuroscience*, *8*, 407-416.

Utama, N. P., Takemoto, A., Koike, Y., & Nakamura, K. (2009). Phased processing of facial emotion: An ERP study. *Neuroscience research*, *64*, 30-40.

Van't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, *108*, 796-803.

Vidulich, R. N., & Krevanick, F. W. (1966). Racial attitudes and emotional response to visual representations of the Negro. *The Journal of social psychology*, *68*, 85-93.

Vollm, B. A., Taylor, A. N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., ... & Elliott, R. (2006). Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *Neuroimage*, *29*, 90-98.

Vuilleumier, P., Armony, J.L., Driver, J., & Dolan, R.J. (2001). Effects of attention and emotion on face processing in the human brain. An event-related fMRI study. Neuron 30, 829-841.

Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., & Bara, B. G. (2004). Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *Journal of cognitive neuroscience*, *16*, 1854-1863.

Westie, F. R., & De Fleur, M. L. (1959). Autonomic responses and their relationship to race attitudes. *The Journal of Abnormal and Social Psychology*, *58*, 340-347.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expression modulate amygdala activity without explicit knowledge. *Journal of Neuroscience*, *18*, 411-418.

Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly*, *59*, 189-202.

Willis, J., & Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychoogical Science.* 17, 592-98.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103-128.

Winston, J., O'Doherty, J., & Dolan, R. J. (2003). Common and distinct neural responses during direct and incidental processing of multiple facial emotions. *NeuroImage*, *20*, 84-97.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces.*Nature neuroscience*, *5*, 277-283.

Woodworth, R.S. (1938). *Experimental psychology*. Oxford, England: Holt.

Yang, T. T., Menon, V., Eliez, S., Blasey, C., White, C. D., Reid, A. J., Gotlib, I. H., & Reiss, A. L. (2002). Amygdalar activation associated with positive and negative facial expressions. *NeuroReport*, *13*, 1737-1741.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of experimental psychology*, *81*, 141-145.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*, 747-759.

Yun, K. Watanabe, K., & Shimojo, S. (2012). Interpersonal body and neural synchronization as a marker of implicit social interaction. *Scientific Reports, 2*, 959.

CAPÍTULO 2

OBJETIVOS Y JUSTIFICACIÓN DE LA INVESTIGACIÓN

Como seres sociales que somos, a lo largo de la vida no dejamos de interaccionar con los demás y aprender de las consecuencias de nuestros actos en esas interacciones, para así ser capaces de adaptar nuestra toma de decisiones en el futuro. En este trabajo de investigación hemos abordado, desde la perspectiva de la neurociencia cognitiva social, el estudio de estos procesos de interacción social. Hemos investigado el procesamiento de información clave de los rostros de las personas que nos sirve para regular la toma de decisiones. Asimismo, analizamos la forma como usamos la información de las consecuencias de nuestras decisiones para regular interacciones futuras, todo ello mediante la exploración de ambos procesos usando un mismo paradigma y una misma metodología.

Como hemos expuesto en la introducción, los encuentros cara a cara con otras personas hacen posible la percepción de cierta información que nos permite intuir sus estados de ánimo y lo que probablemente harán a continuación, o hacernos una idea de lo que pueden estar pensando. Para ello hacemos uso del conocimiento previo que tenemos de dicha persona (a nivel individual), además de cierta información acerca del estado actual de esa persona, que deducimos de algunos elementos presentes en el rostro del individuo.

Pero en ausencia de conocimiento previo acerca de la persona en cuestión, hacemos inferencias individuales y generales que nos permiten regular nuestro comportamiento con esa persona. Para ello usamos, por un lado, información que extraemos de aspectos variantes de su rostro, como son las expresiones faciales, que procesamos y vamos actualizando de una forma muy automática. Y, por otro lado, hacemos uso del conocimiento que tenemos respecto a las posibles categorías sociales (a nivel social o grupal) en las que se podría ubicar al individuo. Esta categorización social la realizamos principalmente a partir de la percepción automática de aspectos invariantes del rostro, como la edad de la persona, su raza, o su sexo. En la toma de decisiones tenemos en cuenta ambos tipos de claves sociales, variantes e invariantes, para extraer información de los rostros y actuar de forma acorde a lo que nos transmiten. Con todo ello generamos creencias y pautas de comportamiento con los demás. Esto suele ocurrir de manera rápida y automática, es decir, sin que nos demos cuenta de ello, y depende en gran medida de nuestra historia de aprendizaje personal y bagaje cultural.

En este contexto, y de una forma más concreta, el objetivo principal de esta tesis doctoral es estudiar cómo influyen en las interacciones con los demás, en situaciones de toma de decisiones de cooperación, ciertos factores sociales que extraemos a partir de información presente en los rostros de las personas con quien interactuamos. En este análisis distinguiremos entre información invariante, como la raza, y variante, como la expresión emocional. La serie de estudios presentados a continuación se centran en explorar cómo estos factores modulan y sesgan nuestras interacciones sociales empleando un juego económico denominado “Juego de Confianza”.

En la adaptación que hemos hecho de este paradigma para conseguir nuestros objetivos hay dos jugadores. Los participantes en nuestros estudios veían fotografías de las caras de diferentes personas en el ordenador, de raza blanca o negra, y con expresiones de enfado o alegría. Estas personas eran sus hipotéticos compañeros de juego, con quienes debían decidir si cooperar o no, compartiendo una cantidad del dinero en juego. En cada ensayo, el participante comienza con una cantidad de dinero y tiene que decidir si lo comparte o no con su pareja de juego. Si se queda con el dinero (no coopera) la pareja de juego se queda sin nada; si por el contrario decide cooperar y comparte su dinero, el otro jugador recibe la cantidad compartida multiplicada por x. Luego supuestamente esta pareja decide si ser recíproco con el participante, en cuyo caso la suma se divide entre los dos, o si se queda con todo, con lo cual el participante lo pierde todo.

Con este paradigma podemos estudiar la influencia de cierta información presente en los rostros de los compañeros, en las decisiones de cooperar de los participantes. Además, permite estudiar otro aspecto esencial de las interacciones sociales como es la respuesta de la otra persona como consecuencia de nuestra decisión (es decir, el ser recíproco o no). Esto tiene unas connotaciones motivacionales importantes, y nos sirve de feedback para regular nuestras respuestas futuras. En el contexto del Juego de Confianza, las consecuencias de confiar en una persona y cooperar con ella pueden ser positivas, si esa persona es recíproca y coopera con nosotros, o negativas si no coopera con nosotros y se queda con todo el dinero. Tener en cuenta esa información es crucial para regular nuestro comportamiento futuro con esa persona. En este sentido, como segundo objetivo de esta tesis, estamos interesados en explorar cómo se procesa el feedback de cooperación vs. no-cooperación en función de que sea proporcionado por personas con distinta identidad, emoción o raza, en el contexto de la toma de decisiones que se lleva a cabo en los Juegos de Confianza.

En este contexto, y de forma más concreta, en la presente tesis doctoral se plantean los siguientes objetivos:

1) Estudiar cómo influyen a nivel comportamental la expresión emocional y la raza de las personas con quien interactuamos, en nuestro comportamiento de cooperación con ellos. Estudiaremos el papel de estas claves, por un lado, cuando no son predictivas de reciprocidad o no-reciprocidad por parte del compañero, y por otro lado, cuando se establecen contingencias reales de cooperación con la emoción, bien de manera que la emoción prediga sus consecuencias naturalmente asociadas (esto es, expresión de alegría predice mayor cooperación que expresión de enfado), o bien asociándolas de una manera contraria a lo esperado (expresión de enfado predice cooperación).

2) Explorar los mecanismos cerebrales implicados en el procesamiento de estas claves (raza, y expresiones emocionales) durante las interacciones sociales. De forma específica, nuestro objetivo era observar cómo el cerebro de los participantes procesaba la raza y la emoción de las caras mientras estos participaban en un Juego de Confianza; nos centramos en el análisis del llamado potencial N170, por ser un correlato cerebral que refleja el procesamiento de caras.

3) Estudiar cómo al introducir la identidad como un factor predictivo de cooperación, cambia la influencia de la expresión emocional en las decisiones de cooperación, explorando los cambios tanto a nivel comportamental como a nivel electrofisiológico del procesamiento de rostros cooperativos vs. no cooperativos.

4) Por otro lado, nos centramos en el estudio del procesamiento del feedback obtenido de esas interacciones. Tomando la FRN como un índice del procesamiento del feedback y teniendo en cuenta estudios previos, exploramos las modulaciones de dicho índice por la emoción (y la raza) de la persona de la cual se obtiene un determinado resultado, que puede ser de pérdida o de ganancia. Igualmente estudiaremos los cambios que se producen al hacer la identidad predictiva de la tasa de reciprocidad, haciendo a unas parejas de juego más “confiables” que a otras.

Nos centramos por un lado en medidas comportamentales de tasas de cooperación, y por otro, mediante un registro electroencefalográfico de alta densidad (HDEEG), que se basa en el registro de la actividad eléctrica cerebral, analizamos el curso temporal de potenciales evocados asociados al procesamiento de raza y emoción. Analizamos también los índices electrofisiológicos del procesamiento del resultado de la cooperación, es decir, del feedback que reciben los participantes en relación a si su compañero ha decidido cooperar con ellos o no. En estos análisis tendremos en cuenta el tipo de persona de la cual proviene supuestamente la información del feedback, y la expresión emocional de su rostro.

Esta técnica ofrece índices indirectos, algunos de ellos bien documentados en la literatura, del proceso a estudiar que nos concierne, como son el potencial N170, sensible a las caras, y la FRN, señal sensible al feedback obtenido. Si bien la sensibilidad funcional del N170 ante la emoción y la raza se viene estudiando desde las últimas décadas, la presente investigación pretende ofrecer una aportación al tratar de estudiarlos conjuntamente y desde un paradigma "social" que, a pesar de las limitaciones que conlleva el que se haga desde el ámbito del laboratorio, puede aportar pistas sobre lo que subyace a nivel cognitivo en las interacciones reales con los demás. Asimismo, en el caso de la negatividad asociada al feedback, aunque se vienen usando paradigmas dentro de la toma de decisiones y los juegos económicos para investigar cómo afectan las expectativas y otros aspectos de tipo motivacional, hasta la fecha escasa investigación hace referencia a cómo puede este correlato verse modulado por claves sociales y faciales de los demás en un contexto de interacción social.

En la primera serie experimental (Capítulo 3) se llevaron a cabo tres experimentos en los que se estudia, a lo largo de un Juego de Confianza, cómo la expresión emocional del rostro del compañero de juego influye sobre la toma de decisiones en distintas condiciones: 1- cuando la expresión emocional no es predictiva de la tasa de cooperación (Experimento 1); 2- cuando de forma explícita se informa que no está ligada a un significado social (Experimento 2); y 3- cuando se establecen asociaciones reales entre la expresión emocional y sus consecuencias (congruentes o incongruentes con la emoción; experimento 3). Para ello usamos medidas comportamentales de tasa de cooperación de los participantes, a partir de sus respuestas de cooperación o no cooperación en cada ensayo.

La segunda serie experimental (Capítulo 4) consta de una serie de dos experimentos en los que se lleva a cabo un Juego de Confianza, manipulando la emoción y raza de las parejas de juego. Por un lado exploramos de nuevo a nivel comportamental cómo modulan estas claves las respuestas de cooperación de los participantes. Por otro lado, mediante medidas electrofisiológicas, exploramos los correlatos neurales asociados a las caras desde el momento de su presentación. En concreto, nos centramos en la exploración del N170, que por su relevancia en la literatura del procesamiento facial, puede proporcionar información sobre cómo interactúan estas claves sociales, cuando son procesadas conjuntamente y aparecen de manera incidental (sin que haya contingencias con el comportamiento de los demás). El segundo experimento replica el primero, igualando la muestra en sexo, por ser esta una variable importante a controlar ya que puede haber diferencias entre sexos cuando se tratan de decisiones económicas y de confianza. También se modifica la cantidad a compartir con motivos prácticos de simplificación.

En la tercera serie (Capítulo 5), llevamos a cabo un tercer estudio estableciendo contingencias de cooperación entre las diferentes parejas de juego y sus tasas de cooperación, con el fin de determinar el efecto de la emoción sobre las tasas de cooperación cuando entra en juego un predictor más fiable como es la identidad de cada persona. Manipulamos pues esta variable de manera que algunas parejas de juego eran más propensas a cooperar, mientras que otras eran más propensas a no cooperar. Asimismo, retomamos los dos experimentos anteriores (capítulo 4) y analizamos y ponemos en relación la negatividad asociada al feedback (tomando como índice cerebral la FRN), teniendo en cuenta la expresión emocional de las parejas de juego, y el tipo de pareja del que se trate, más cooperativa o menos cooperativa.

Hemos organizado estas series experimentales en tres capítulos separados, cada cual con su propia introducción y discusión. El Capítulo 3 y 4 son artículos que ya han sido publicados en diferentes revistas, y el Capítulo 5 está en preparación para enviar próximamente. A lo largo de los siguientes capítulos expondremos con detalle cada serie experimental, los resultados comportamentales y de EEG obtenidos. Concluiremos en el último capítulo con una discusión general en el que se integrarán los hallazgos más relevantes y se discutirán sus implicaciones en el ámbito de la neurociencia cognitiva social, así como los procesos automáticos y no conscientes que se dan en encuentros con gente de otra raza y que nos provocan a veces reacciones de prejuicio, o conductas de aproximación–rechazo ante ciertas expresiones faciales emocionales. Para finalizar, haremos mención a investigaciones en curso y futuras que se plantean en este ámbito.

CAPÍTULO 3

INTERPERSONAL EFFECTS OF EMOTION IN A MULTI-ROUND TRUST GAME

(Efectos interpersonales de la emoción en un Juego de Confianza)

The content of this chapter has been published as Tortosa, M. I., Strizhko, T., Capizzi, M., & Ruz, M. (2013). Interpersonal effects of emotion in a multi--round Trust Game. *Psicológica*, *34*, 179-198.

ABSTRACT

Emotions displayed by others are pivotal ingredients of the decisions we make in social contexts. However, most of the research to date has focused on the subjective emotion of the decider rather than on the emotional expressions of the partners in the interaction. The present investigation was designed to explore how happy and angry facial expressions modulate cooperative responses in multi-round Trust Games. Our results show that happy partners generate higher levels of trust than angry partners even after repeated experience in a context in which emotional displays are not predictive of the partners' cooperation rates. This effect disappears once the social meaning of emotional displays is eliminated from the game. An additional study shows that participants are able to learn specific associations between discrete emotions and positive or negative cooperative tendencies, although they need more evidence when the associations counteract prior expectations. Overall, our results stress the reliability of discrete emotions as cues in interpersonal interactions and the resilience of the effect of these positive and negative cues in contexts in which they lack real predictive power.

Resumen

Las emociones que otras personas expresan juegan un papel importante en las decisiones que tomamos en contextos sociales. Sin embargo, la mayoría de la investigación hasta la fecha se ha focalizado en la emoción subjetiva de la persona que toma la decisión, en vez de en la emoción mostrada por los compañeros en una interacción. Nuestro estudio se diseñó para explorar cómo las expresiones de felicidad y enfado de otras personas afectan a las respuestas de cooperación en un Juego de Confianza de interacciones multiples. Los resultados muestran que los compañeros felices generan niveles de cooperación más altos que los enfadados, incluso después de interacciones repetidas en las que las emociones no predicen la tasa de cooperación de los compañeros. Dicho efecto desaparece cuando el significado social de las emociones se elimina del juego. Otro experimento adicional muestra que los participantes son capaces de aprender asociaciones específicas entre emociones discretas y diferentes tendencias cooperativas, aunque necesitan más evidencia cuando la asociación es contraria a las expectativas previas. En conjunto, nuestros resultados muestran que las emociones se emplean como señales en las interacciones entre personas, y que su efecto es duradero incluso en contextos en los que carecen de predictividad real.

INTRODUCTION

In our daily lives, we navigate through elaborate, dynamic systems of social interactions. Decision-making is guided not only by our knowledge of the purpose and rules of the interactions, but also by our beliefs and expectations about those with whom we interact. Successful interactions with other people therefore demand the ability to attribute to others internal mental states such as intentions, beliefs, feelings, and goals (this ability is also called theory of mind, ToM; Frith & Frith, 2003). We use these assumed states to explain and predict others' behaviors (Frith & Frith, 2006).

While we are motivated to form accurate impressions of those who can influence the outcomes of social interactions (Vonk, 1998), our envisioning of their mindset can be modulated by various factors. For example, when we enter a social interaction with limited knowledge about those with whom we interact, we may rely on contextual cues in the qualities we perceive in them (their gender, race, social status, facial expressions, attractiveness and so forth; see, for example, Delgado, Frank, & Phelps, 2005; Ruz, Moser & Webster, 2011). Strikingly, many of these factors affect us automatically, sometimes even without conscious awareness (Bargh, Chen, & Burrows, 1996; Bargh & Ferguson, 2000), and lead to expectations about how we and others may act (Olson, Roese, & Zanna, 1996).

One important factor influencing social decision-making is emotion (Adolphs, 2003; Olsson & Ochsner, 2008). For example, the induction of positive or negative emotional states on the decider respectively increases or decreases trusting behaviors (Dunn & Schweitzer, 2005; see also Harle & Sanfey, 2007). Most of research to date, however, has focused on the *intrapersonal* effects of emotion, or how moods or emotional states of the decider affect judgment and decision-making in several situations (see Angie, Connelly, Waples & Kligyte, 2011, for a recent review). However, the interpersonal effects of emotions, or how the emotions expressed by other people could be used as cues to predict their most likely behavior, have been much less explored (Van Kleef, de Dreu, & Manstead, 2010).

Given their functional role in communicating intentions (Darwin, 1872; Fridlund, 1995; Keltner & Haidt, 1999), facial expressions of emotion are especially strong candidates to influence trust decisions in social encounters, as they play a major role in indicating when a person is willing to be cooperative and trustworthy and when a person is not (Buck, 1984; Boone and Buck, 2003). Along evolution, we have learned to associate different emotions with specific meanings. In most cases, positive emotions such as happiness predict positive consequences whereas negative emotions such as anger indicate that bad things may happen (Darwin, 1872; see Ruz & Tudela, 2011; Ruz, Madrid & Tudela, 2012). These associations, crafted along the years, help us to adjust our behavior in light of the predispositions of others.

The link between emotions and social decision-making also emerges in research conducted in the field of social neuroeconomics (Fehr & Camerer, 2007; Sanfey, 2007), which widely employs experimental economic games to study the patterns of social behaviors in interactive situations. This research indicates that decision makers do not always behave "rationally", or follow the strategy of strict self-interest and individual maximization (Camerer, 2003). The experimental Trust Game (Berg, Dickhaut, & McCabe, 1995; Camerer & Weigelt, 1988) is a consistent generator of "irrational" decisions. The game involves at minimum two players, a *trustor* and a *trustee*. The trustor is endowed with a sum of money and has to decide whether or not to share it with her game partner. If she keeps the money for herself, the trustee gets nothing. If she decides to share, the trustee receives the initial endowment multiplied by an amount. If he then reciprocates the trust, the sum is divided between the two players; otherwise the trustor obtains nothing. In this game, the decision of the trustor is hazardous because the trustee's reciprocation is not enforced by the rules. Therefore, the individually rational strategies in single-round games are not cooperating for the trustor and not reciprocating for the trustee. Still, substantial amounts of trust are observed across studies (Berg et al., 1995), which are attributed to altruism and reciprocation that activate reward brain circuits (Fehr & Camerer, 2007).

Social preferences for trust are not unconditional but depend on the belief that the partner is likely to reciprocate the trust (Camerer, 2003), combined with the general tendency of people to trust others (Berg et al., 1995). Two different studies have shown that happy facial expressions – either schematic line drawings (Eckel & Wilson, 2003) or photographs of people (Scharlemann, Eckel, Kacelnik, & Wilson, 2001; see also Averbeck & Duchaine, 2009) - generate higher levels of *initial* trust in one-shot games. Moreover, the choices seem to be influenced by the facial dynamics that distinguish between genuine and fake smiles (Krumhuber, Manstead, Cosker, Marshall, Rosin, & Kappas, 2007; Niedenthal et al., 2010), with authentic smiles generating higher cooperation rates.

The studies that have explored the evolution of trust in an iterated exchange presented either no photos, and thus no information regarding the partners' displayed emotions (King-Casas, Tomlin, Anen, Camerer, Quartz, & Montague, 2005), or 'expressionless' neutral photos of the game partners (Delgado, Frank, & Phelps, 2005). Other recent studies that have used multi-round exchanges have focused on the effects of negative emotion as anger on future interactions in computer-mediated negotiations (Van Kleef et al. 2010) and disappointment vs. anger with the tit-for-tat strategy (Wubben, De Cremer, & van Dijk, 2009). They did not present faces though, but statements expressing the emotions in question.

Our study explored how the behavior elicited by emotional facial expressions of happiness and anger is maintained over the course of an extended social interaction. Along three experiments, we investigated: (1) whether happy and angry facial expressions of partners in a Trust Game modulate trust decisions even after several rounds in which such emotional expressions are not predictive of the partners' cooperation rates; (2) whether these long-lasting effects remain once emotional displays are not linked to the partner in the game; and (3) whether people are able to use happy and angry facial expressions as cues that predict their natural or their unnatural consequences in inter-personal situations (i.e. non-cooperative and cooperative tendencies, respectively). Our hypotheses predict: (1) a higher rate of cooperative responses after happy than after angry emotional expressions; (2) no effect of non-social emotions; and (3) a rapid association of positive and negative emotions with cooperative and non-cooperative purpose/intent, while a delayed learning of associations that are not consistent with initial priors linking emotions and their most likely consequences.

Experiment 1

METHOD

Participants

32 students (8 males, mean age of 21 years) from the University of Oxford (12) or the University of Granada (18) participated in exchange for course credits. They all signed a consent form approved by the local Ethics committees and received a chocolate token for their participation. Two of them were excluded from analyses because did not have enough observations in all conditions.

Stimuli and procedure

At the beginning the session, participants were instructed that the experiment explored the cooperation patterns that emerge between people during the so-called Trust Game. The participant would play multiple rounds with three different players over the course of the game. At the beginning of every round, participants were presented with a symbolic Pound/Euro and had to decide whether to keep it (by pressing the *k* on the keyboard) or share it with their partner (by pressing *s* on the keyboard). The keep decision would yield no earnings for the partner and end the trial. The share decision would result in £5/5 Euro given to the partner who, in turn, would decide whether (1) to reciprocate the cooperation, in which case each of them would receive £2.5/2.5 Euro; or (2) not to reciprocate, in which case the participant would receive nothing because the partner kept the £5/5 Eur. The participants' goal was to maximize their payoffs in the game, and they were told that *mutual* cooperation was the best strategy for reaching this goal. They could, however, to the best of their knowledge freely decide on every trial whether to trust or not their partner. The prize for maximizing their payoffs was a chocolate bar once the game was finished. Participants were also told that even though their partners were represented by photos on the computer, their behavior mimicked normal patterns of play by real people. Therefore, they did know that they were not playing against real people on-line, but the instructions stressed that the face photos represented the choices of normal partners. Participants were not told about the different emotions that partners would be displaying, and thus they were unaware of the main goal of the study, which was to explore how emotions influenced cooperation rates. At the end of the game, all participants received a chocolate bar.

The game was presented on a PC running E-Prime software (Schneider, Eschman, & u ccolotto, 2002). Frontal photographs of three female (see Aguiar, Brañas-Garza, Cobo-Reyes, Jimenez & Miller, 2009) white faces displaying happy, neutral, or angry emotional expressions were selected from the NimStim face stimulus set (Tottenham, Tanaka, Leon, McCarry, Nurse, Hare et al., 2009) to represent the participants' partners. All stimuli were presented against a grey background (see Fig. 3). Every trial began with a 1-sec presentation of a Pound or Euro symbol, replaced by a fixation point (+) for another 1 sec, and was followed by the picture of the partner for that trial. After 1.5 sec, the picture was replaced by the fixation point for 1 sec. Next, a question mark (?) was displayed for 2.5 sec; this served as a prompt for the participants to make their decision for that trial. Finally, the participants received feedback about the payoffs in the trial (see Fig. 3), which was displayed for 2 sec. If they decided to keep the money, the message was 'You have decided not to cooperate. You add £1/1 Euro and your partner adds £0/0 Euro'. If they decided to share, either 'You have decided to cooperate. Your partner receives £5/Eur and decides to correspond. You add £2.5/2.5 Euro and your partner adds £2.5/2.5 Euro' or 'You have decided to cooperate. Your partner receives £5/5 Euro and decides not to correspond. You add £0/0 Euro and your partner adds £5/5 Euro' were presented. Participants played this game 84 times with each of the three partners (for a total of 252 trials), who on a trial-by-trial basis displayed random happy, neutral, or angry emotions with equal probability and reciprocated at the constant rate of 50% regardless of their emotional expression. The session lasted about 30 minutes.

We analyzed the percentage of participants' cooperation rates across conditions. Our main variable was the emotion displayed by the partner, which was manipulated at 3 levels: happy, neutral, and angry. In addition, we included two more variables in the design. First, responses were divided in blocks of 50 trials to examine the effect of practice with the task. Second, we included the feedback that participants received from their partner in the previous trial to explore how this may have affected their subsequent decision (3: non-cooperation, non-reciprocated cooperation and reciprocated cooperation). Thus, the average levels of acceptance rates per participant and condition were analyzed with repeated-measures ANOVAS with the factors Emotion (3) x Block (5) x Feedback (3).

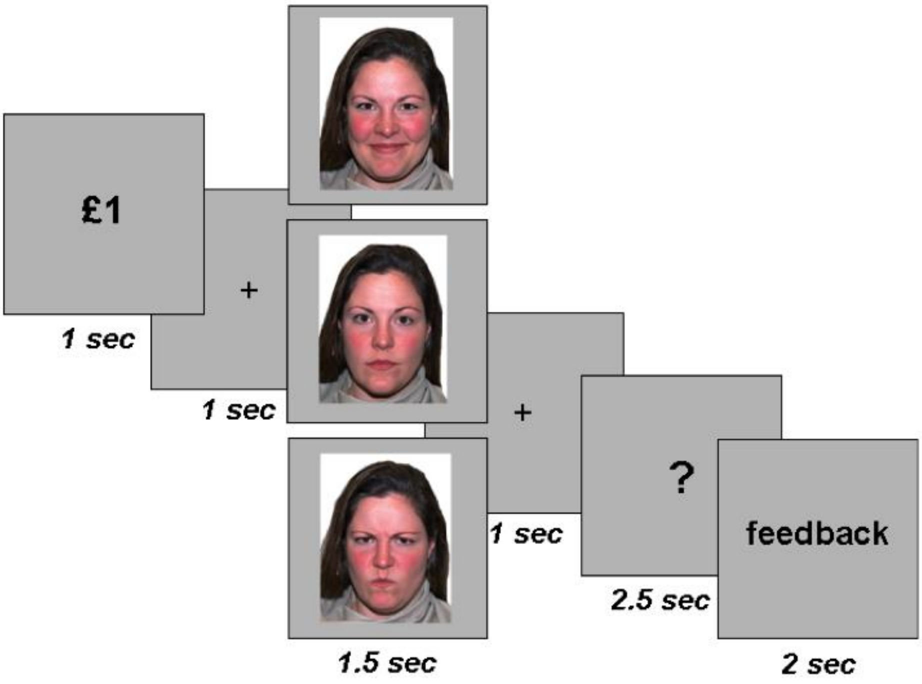


Figure 3. Sequence of events in a trial.

RESULTS

Participants cooperated on 62% of the trials (SD=33.7). Results showed a main effect of Emotion, $F_{2,58}$=12.53, p<0.001, as participants cooperated more with happy than with angry partners (70.4% vs. 49.4%), $F_{1,29}$=14.83, p<0.001. The difference between angry and neutral (65.1%) was also significant, $F_{1,29}$=13.19, p=0.001, but between happy and neutral it only reached marginal significance, $F_{1,29}$=3.66, p=0.066 (see Fig. 4). The effect of Block of trials was not significant, F<1. Nevertheless, to test whether the factor Emotion was still significant after repeated experience with the game, we evaluated the effect of the partners' emotional expression in the last block of trials (Block 5). In this block, cooperation rates were still lower for angry (44%) than for happy (67%), $F_{1,29}$=11.8, p=0.001 or neutral partners (64%), $F_{1,29}$=9.41, p<0.01. The difference between happy and neutral partners was not significant, F<1. The variable Feedback did not modulate cooperation rates, F<1, and no interaction reached significance levels (all ps>0.1).

An additional analysis was performed to test for potential differences between participants from Oxford and Granada universities. However, this between-subject variable did not produce any reliable effect (all relevant Fs<1).

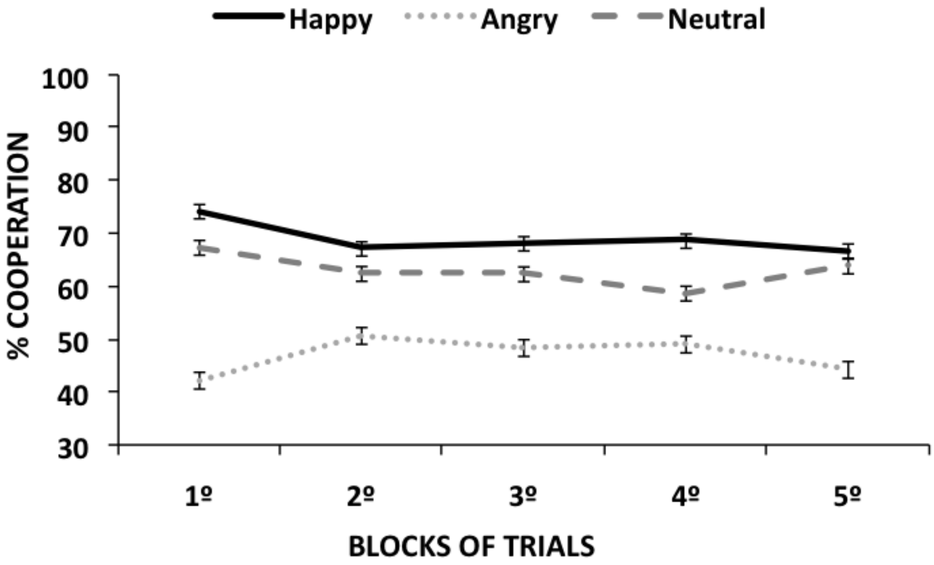


Figure 4. Participants' cooperation rates with partners displaying happy, angry or neutral facial expressions along five blocks of trials in Experiment 1.

DISCUSSION

In Experiment 1, participants cooperated more with happy partners than with those displaying neutral or angry facial expressions. This result is fits with previous findings that relate happiness expressions with trustworthiness (Eckel & Wilson, 2003) or signals of cooperative intents (Fridlund, 1995). In addition, it extends the results of experiments that showed that happy expressions relate to initial levels of high cooperation rates in single-round Trust Games (Eckel & Wilson, 2003; Scharlemann et al., 2001). The lack of interaction between the factors of emotion and block of trials, and the significant effect of emotional expressions on participants' cooperation rates during the last block of trials, strongly suggests that this emotion-related bias persists even after repeated experience in a game setting in which emotional expressions are not predictive of the partner's reciprocation of cooperation. In addition, the feedback that participants received on the previous trial seemed to have no effect on subsequent trust decisions, as cooperation rates were not altered across feedback conditions.

However, at this point there is a potential and less appealing alternative explanation to our results. The effect that emotional expressions had on the cooperation rates of participants may derive from general mood effects or affective priming (i.e. cooperation may be enhanced by the mere display of positive stimuli) rather than by the social meaning of emotions. From this perspective, the presentation of any emotional material unrelated to social interactions may have produced the same results due to their condition of primes with evaluative value.

To test this alternative hypothesis, we set up Experiment 2. Instead of introducing additional confounds by changing the nature of the affective stimuli, we repeated the exact same procedure using emotional photos of the partners, but modified the instructions that participants received to devoid the emotional expressions of their social meaning.

Experiment 2

METHOD

Participants

32 students (1 male, mean age of 21.9 years) from the University of Granada participated in exchange for course credits. They all had normal or corrected-to-normal vision.

Stimuli and procedure

The experiment was the same as the previous one, except for the modified instructions. The background story remained the same. They were going to play a Trust Game with several different partners, who were represented by photos on the computer but had a behavior that mimicked normal patterns of play by real people. Their goal, again, was to maximize pay-offs and they were told that *mutual* cooperation was the best strategy for reaching this goal. However, whereas participants in Experiment 1 were not told about the emotions that the partners would display, in the current one they were informed that the computer *randomly* assigned different emotions to the partners. As in the previous experiment, though, they were reminded that the identity of the partner was relevant for the game.

The design included the factors Emotion (Happy, Angry, Neutral), Block (5) and Feedback (noncooperation, non-reciprocated cooperation and reciprocated cooperation) as within-subject variables.

RESULTS

Participants cooperated on 61.9% of the trials (SD=12.8). There was a main effect of the feedback, $F_{2,34}$=4.79, p=0.01, as participants tended to cooperate more after not cooperating in the previous trial compared to when the feedback was both a non-reciprocated, $F_{1,17}$=4.77, p<0.05, or a reciprocated cooperation, $F_{1,17}$=8.26, p=0.01. In contrast to the previous experiment, there was no effect of emotion, F<1. No interaction reached significance levels (all ps>0.1).

DISCUSSION

Results from Experiment 2 indicate that the biasing effect of emotion on cooperation rates in a Trust Game was not due to automatic mood or affective priming effects. If this was the case, we should have found the same effects in Experiment 2, in which everything remained the same except for the association of the partners to their emotional expression in the context of a social interaction. Instead, in the current experiment we found that the cooperation rate of participants was equal across happy, neutral and angry partners. This suggests that the link between the partners and their expression in a social context drove the effect of emotions observed in Experiment 1, and not mere mood or affective priming effects.

Together, experiments 1 and 2 suggest that the priors that we hold relating positive emotional expressions with cooperative consequences are strong (see Averbeck & Duchaine, 2009; Ruz & Tudela, 2011) and resistant to evidence that regards them as not informative (as emotional expressions did not predict the partners' cooperation intents). If this were so, we would expect that in game context in which emotional displays predicted the cooperative consequences that they are naturally associated to, participants would adjust their cooperative behavior in a fast way. On the contrary, if the game context associated emotions with the opposite of their natural associations, people would need more evidence (i.e. trials) to adjust their behavior. These were the manipulations in Experiments 3.

Experiment 3

METHOD

Participants

26 students from the University of Oxford (13 females, mean age 21.5) participated in the experiments. They all signed a consent form approved by the local Ethics Committee.

Stimuli and procedure

The current experiment followed the same structure as Experiment 1 but added a contingency between facial expressions and reciprocity rates. For one group of participants (bias-consistent), emotions in the game were associated to their natural consequences. All partners displayed happy, angry and neutral expressions with equal probability (33.3%), but their reciprocation rate depended on their emotional display. Each reciprocated on 80% of the trials in which their facial display was happy, on 20% when it was angry, and on 50% when they had a neutral expression. For another group of participants (bias-inconsistent), the contingency between facial expressions and reciprocity rates were reversed, and now the emotion displayed by the partner predicted the opposite of their natural consequences. Thus, partners reciprocated in 20% of the trials when their facial expression was happy, 80% when it was angry, and 50% when they had a neutral display. Only partners' expressions and not identities were predictive of their trustworthiness. The design was multifactorial with Emotion (Happy, Angry, Neutral), Block (5) and Feedback (non-cooperation, non-reciprocated cooperation and reciprocated cooperation) as within-subject variables, and Group (bias-consistent, bias-inconsistent) as between-subject factor.

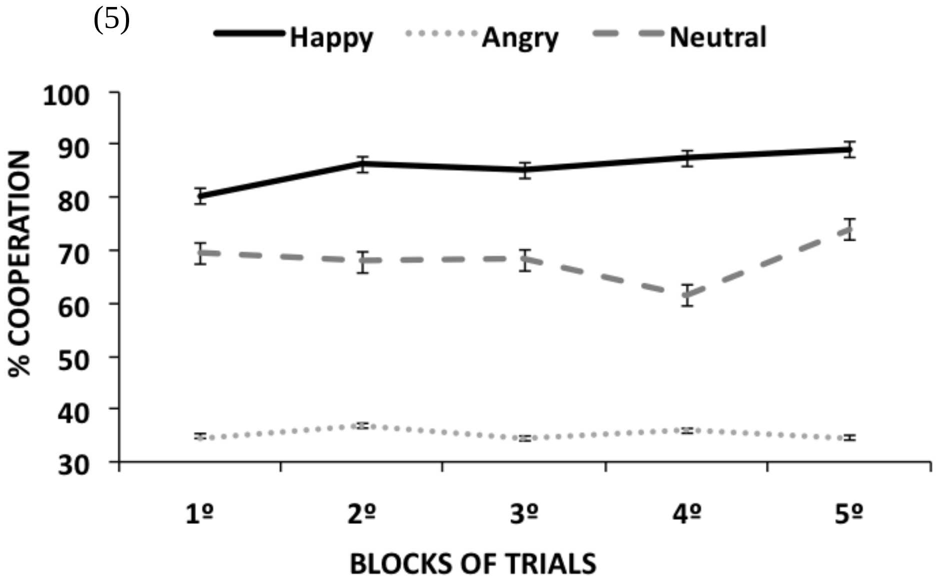
RESULTS

Mean cooperation rate was 63.7% (SD=33.2). The ANOVA showed an interaction between Experimental group, Emotion, and Block, $F_{8,192}$=5.34, p<0.001, due to differences between groups in cooperation rates across the blocks.

In the bias-consistent group, where contingencies between emotion and cooperation rates were the expected, there was a main effect of Emotion, $F_{2,24}$=31.85, p<0.001.

Participants cooperated more with happy (85.7%) than with neutral (67.9%), $F_{1,12}$=18.47, p=0.001 or angry partners (35.9%), $F_{1,12}$=40.42, p<0.001 (see Fig. 5). The difference between neutral and angry partners was also significant, $F_{1,12}$=24.65, p<0.001. Crucially, there was no interaction with Block, F<1, as the effect of Emotion was constant along the task. In this group, the variable Feedback modulated cooperation rates, $F_{2,24}$=4.25, p<0.05. Participants cooperated more after a reciprocated cooperation feedback (68.2%) than after a non-reciprocated one (58.9%), $F_{1,12}$=5.49, p<0.05, and more than after a non-cooperative decision (63.2%), $F_{1,12}$=5.75, p<0.05. There was no difference in participants´ cooperative behavior between non-cooperation and non-reciprocated feedbacks, $F_{1,12}$= 1.12, p>0.3.

In the bias-inconsistent group, where contingencies were reversed, there was a main effect of Emotion, $F_{2,24}$=5.00, p<0.05. Overall, cooperation rates were higher for angry (76.8%) than for happy (51.0%) partners, $F_{1,12}$=6.38, p<0.05, and higher for neutral (65.4%) than for happy partners, $F_{1,12}$=6.05, p<0.05, with non-significant differences between angry and neutral, $F_{1,12}$=2.12, p>0.1 There was also an interaction between Emotion and Block, $F_{8,96}$=7.37, p<0.001, which showed that the effect of Emotion appeared as participants acquired practice with the game contingencies. In the first block, cooperation rates tended to be equal for happy (65.9%) angry (61.4%) and neutral partners (67.5%), all Fs<1. Block 2 showed a main effect of Emotion, $F_{2,24}$=3.53, p<0.05, that increased until Block 5, $F_{2,24}$=22.91, p<0.001. In this last block, cooperation rates were higher for angry (93.1%) than for happy (43.8%), $F_{1,12}$=42.53, p<0.001, and neutral partners (56.6%), $F_{1,12}$=27.21, p<0.001, although there were no differences between neutral and happy partners, $F_{1,12}$=2.52, p>0.1 (see Fig.6). In this bias-inconsistent group the variable Feedback did not modulate cooperation rates, F<1.



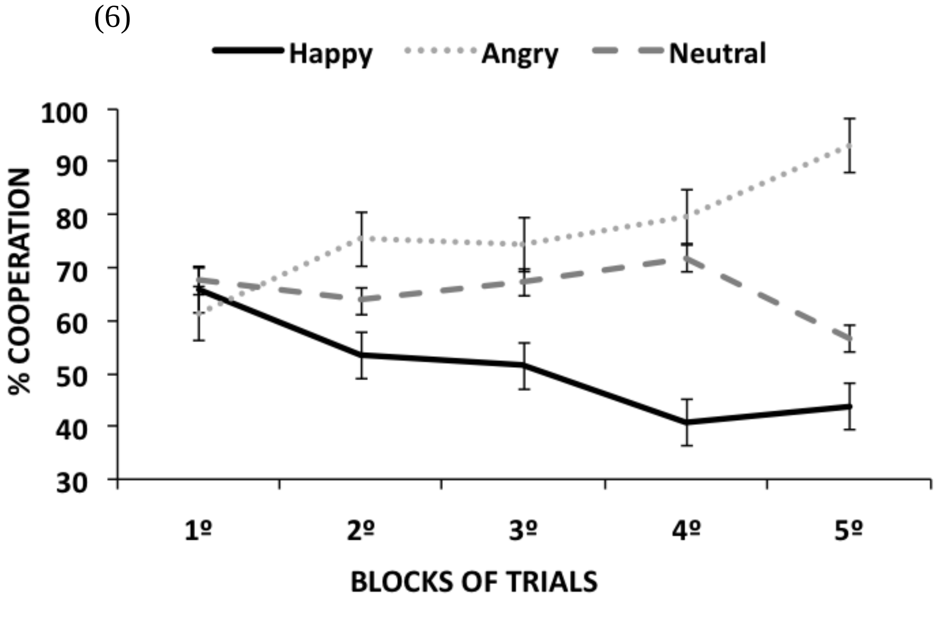


Figure 5 & 6. Participants' cooperation rates with partners displaying happy, angry or neutral facial expressions along five blocks of trials in Experiment 3.

DISCUSSION

Participants learned the contingency between a facial expression and reciprocity rate, as their cooperation mimicked the rates assigned to each type of emotions. They were able to learn such associations when they were consistent with the expectations set by the emotional expressions, but also when they were inconsistent with the links that we hold naturally between valence of emotions and trustworthiness. Participants from the bias-inconsistent group, however, arrived at the 'correct' contingencies later than those in the bias-consistent group, as evidenced by the significant interaction between Group, Emotion and Block. Thus, whereas there were no discrepancies between the expectations generated by the emotional expressions and the behavior of the partners in the bias-consistent group, the initial predictions failed in the bias-inconsistent group, and thus they needed more time to learn the correct cooperation rates.

In addition, and in contrast to Experiment 1, participants' trust decisions were influenced by the previous cooperation feedback from their partners, but only in the bias-consistent group, as they cooperated more after their partners in the previous trial reciprocated the cooperation than when they did not. However, the reciprocity feedback obtained from the previous trial did not modulate cooperation rates in the bias-inconsistent group. We will turn to why this may be so in the General Discussion section.

In sum, these results bear on the bias that people have to hold positive expectations when confronted with an expression of happiness, and negative expectations when confronted with an expression of anger. Most important, our results show that people are able to learn associations between emotions and cooperation tendencies that mismatch their priors (Averbeck & Duchaine, 2009; Ruz & Tudela, 2011) and adapt their behavior to such unnatural associations, even though this takes an extended experience.

GENERAL DISCUSSION

The present study explored the effect of the emotions displayed by partners in a multi-round trust decision-making setting. Experiment 1 revealed that even after prolonged exposition with a game context in which the emotions displayed by the partners were not predictive of their cooperation rates, participants still cooperated more with happy than with angry partners. Experiment 2 supported the idea that this effect was not due to automatic affective biases generated by the mere presentation of emotional stimuli. In Experiment 3, participants quickly adapted their behavior in a game in which emotions predicted their natural consequences and, although it took them longer, they were able to adapt their cooperation rates to the game contingencies when the association between emotion and cooperation rates was counter-intuitive.

Overall, our results are in line with previous research showing that positive and negative emotions modulate initial levels of trust in interpersonal encounters (Eckel & Wilson, 2003), and support models that posit that facial expressions of emotion are salient cues employed to predict the behavior of others in interpersonal social contexts (Van Kleef, de Dreu, & Manstead, 2010). In addition, results from our Experiment 1 extend this literature by showing that the differential levels of trust that happiness and anger generate persist after repeated lack of evidence that emotions are of any use in the game, as they were not associated in any manner to the partners' cooperation rates. Despite this, after more than 200 trials (in the last block of the game), happy partners still generated significantly higher levels of trust than angry ones, and the participants' cooperation rates with partners displaying the latter expression were lower than with those with neutral facial displays. Such bias disappeared completely when the emotions were devoid of their association with the partners by telling participants that they were placed at random by the computer program controlling the game.

In the EASI model proposed by Van Kleef, de Dreu and Manstead (2010; see also Van Kleef, 2009), the context of the social interaction sets the effect that the discrete emotions of others are going to generate on the decision of the perceiver. In general lines, the model posits that in cooperative settings the effect of emotions is funneled through affective reactions, whereas in competitive conditions emotions are employed to draw strategic inferences. The Trust Game we used could be conceived as a cooperative setting, as instructions stressed that *mutual cooperation* between the participant and his/her partners was the best strategy to maximize the payoffs in the game, which was the goal that participants had to fulfill. According to the EASI model, thus, the effects of the partner's emotions would be mainly driven by affective contagion (e.g. Parkinson & Simons, 2009). Happy partners would engender positive feelings in the participants, who would perceive the exchange as safe and would thus feel that their partner would reciprocate their initial trust, whereas angry partners would generate negative emotions which would move the participant away from their partner and thus reduce cooperative intents.

The explanation above does not exclude the possibility that emotional displays could *also* be used strategically in the current game setting. As in Experiment 1 the association between emotion and reciprocation consequences was random, participants may have disregarded the cooperation behavior of their partners to be guided only by their emotional expressions, as suggested by the lack of effect of feedback information in this experiment. Experiment 3, however, show that once happy and angry emotional displays provide information regarding the partners' behavior, participants are able to use these cues strategically to guide their choices and adapt their trust levels accordingly.

Results from Experiment 3 suggest that the initial associations between emotions and their consequences are taken into account in the decision-making process, as evidenced by the slow-down in the group learning bias-inconsistent associations between emotions and cooperation rates. In this line, Ruz & Tudela (2011; see also Ruz, et al., 2012) studies suggest that the natural associations between emotions and their consequences are difficult to override and need of additional conflict detection and cognitive control mechanisms, as evidenced by behavioral interference indices and neural activation in the anterior cingulate cortex and prefrontal cortices, as shown by functional magnetic resonance imaging (fMRI; Ruz & Tudela, 2011). Also, Averbeck & Duchaine (2009) using a non-social reward task, showed that people have a *prior bias* to select happy over angry faces as potential sources of reward. Thus, in our experiments the natural expectations engendered by emotions modulated the amount of evidence needed to associate the emotional expressions with their specific associations in each of the games. Whereas in bias-consistent group of Experiment 3 the effect of emotions maintained along the blocks (as evidenced by the lack of interaction between these two factors), in the bias-inconsistent group the interaction between emotion and block of trials suggests that participant needed prolonged evidence with the game contingencies to be able to grasp that happy emotions predicted lack of cooperation whereas angry expressions led to higher cooperation rates.

One intriguing aspect of the present set of studies is the lack of evidence that participants used the feedback from the previous trial to influence their current decisions. Whereas in bias-consistent group of Experiment 3 we observed an effect of feedback in the expected direction (higher cooperation rates after reciprocated than after non-reciprocated trust), there was no hint of such effects in the bias-inconsistent group or in Experiment 1. One possible explanation for this is that our experiments lacked the common association between partners' identity and cooperation tendencies. That is, in none of the experiments the specific identity of the partner predicted cooperation rates (as all partners cooperated at a constant 50% rate). This may have led participants to disregard the feedback from the previous trial, tied to the partner's identity, as relevant information to guide their judgment in the following trial. In contrast, the structure in which the contingencies between emotion and reciprocation rates followed natural expectations, may have made feedback more salient for participants, which may have led them to include it as a relevant factor in their decision. This would fit with the notion that people tend to take more into account the facts (i.e. the feedback) that support their beliefs (for an effect akin to selective exposure, see Kleinhesselink & Edwards, 1975). Future studies associating different identities to differential cooperation rates, combined with emotional displays, should be conducted to help disentangle this matter.

There are some details of the current experiments that limit the scope of our conclusions. First, due to the limited availability of male students at the time data were collected, most of the participants were women, which calls into question the reliability of our results in a male population. To date, however, we have performed some other experiments using similar emotion manipulations with the same Trust Game including equal number of male and female participants, and the factor of gender has never generated a main effect or interacted with any of the variables in the designs (Tortosa, Lupiañez & Ruz, *2013*).

Second, the experimental setting is rather artificial, which makes the extrapolation to real-life social situations more difficult and qualifies the scope of the conclusions that may be derived. We tried to minimize this problem by stressing that the responses of the partners mimicked behavior in real situations. Future studies, however, could be improved on this respect by using videos instead of static pictures of the partners, and/or using actual people as partners. The use of static pictures, however, allowed better experimental control, which could be beneficial, for example, to aid in the adaptation of the paradigm to future electrophysiological experiments.

A drawback related to the artificial nature of the experiments is the extent to which our results can be extrapolated to actual social interactions in which the emotions of the partners are used to infer their internal states in relation to their future cooperative tendencies. It could be argued that emotional faces should be considered as symbolic primes with evaluative value rather than components of actual social interactions. Results from Experiment 2 showing that the effect is lost once emotions are devoid of their social meaning, however, argue against a pure automatic effect from any evaluative prime. In any case, the resolution of this dichotomy falls outside the current study, and should be aided by the use of more naturalistic settings (with real partners or videos of actors, for example).

Overall, our results stress the strong impact of the emotional displays of the people in shaping our decision tendencies, and the resilience of the effect of these positive and negative cues in contexts in which they lack real predictive power. Future research should be aimed at investigating the effect of stable emotional states of others in trust behaviors, and also at exploring the role of other pieces of social information that may be relevant to trust tendencies, such as the description of the moral characteristics of the partners in interpersonal interactions.

REFERENCES

Adolphs, R. (2003) Cognitive neuroscience of human social behavior. *Nature Reviews Neuroscience, 4*, 165-178.

Aguiar, F., Brañas-Garza, P., Cobo-Reyes, R., Jimenez, N., & Miller, L.M. (2009) Are women expected to be more generous? *Experimental Economics, 12*, 93–98.

Angie, A.D., Connelly, S., Waples, E.P., & Kligyte, V. (2011) The influence of discrete emotions on judgment and decision- making: A meta-analytic review. *Cognition & Emotion 15*, 1-30.

Averbeck, B.B., & Duchaine, B. (2009) Integration of social and utilitarian factors in decision making. *Emotion, 5*, 599-608.

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: on the automaticity of higher mental processes. *Psychological Bulletin, 126*, 925-945.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: direct effects of trait construct and stereotype-activation on action. *Journal of Personality & Social Psychology, 71*, 230-244.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity and social history. *Games and Economic Behavior, 10*, 122-142.

Boone, R. T., & Buck, R. (2003). Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior, 27,* 163-182.

Buck, R. (1984). *The communication of emotion.* New York: Guilford Press.

Camerer, C. F., & Weigelt, K. (1988). Experimental tests of a sequential equilibrium reputation model. *Econometrica, 56*, 1-36.

Camerer, C.F. (2003). *Behavioural game theory: Experiments in strategic interaction*. Princeton: Princeton University Press.

Darwin, C. (1872). *The expression of emotions in man and animals*. London: John Murray.

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience, 8*, 1611-1618.

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: the influence of emotion on trust. *Journal of Personality and Social Psychology, 88*, 736-748.

Eckel, C. C., & Wilson, R. K. (2003). The human face of game theory: Trust and reciprocity in sequential games. In E. Ostrom & J. Walker (Eds.), *Trust and reciprocity: Interdisciplinary lessons from experimental research* (pp. 245-274). New York: Russel Sage Foundation.

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences, 11*, 419-427.

Fridlund, A. J. (1995). *Human facial expression: An evolutionary view.* London: Academic Press.

Frith, C. D., & Frith, U. (2006). How we predict what other people are going to do. *Brain Research, 1079*, 36-46.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 358*, 459-473.

Harlé, K.M., & Sanfey, A.G. (2007). Incidental sadness biases social economic decisions in the Ultimatum Game. *Emotion, 7*, 876-881.

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition & Emotion, 13*, 505-521.

Kleinhesselink, R. R., & Edwards, R. E. (1975). Seeking and avoiding belief-discrepant information as a function of its perceived refutability. *Journal of Personality and Social Psychology, 31,* 787-790.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. Science, 308(5718), 78-83.

Krumhuber, E., Manstead, A. S., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion, 7*, 730-735

Niedenthal P.M., Mermillod, M., Maringer, M., & Hess U. (2010). The simulation of smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral & Brain Sciences*, 33, 417-33

Olson, J. M., Roese, N. J., & Zanna, M. P. (1996). *Expentancies*. New York: Guilford.

Olsson, A., & Ochsner, K.N. (2008). The role of social cognition in emotion. *Trends in Cognitive Sciences, 12*, 65-71.

Parkinson, B., & Simons, G. (2009). Affecting others: social appraisal and emotion contagion in everyday decision-making. *Personality & Social Psychology Bulletin, 35*, 1071-1084.

Ruz, M., & Tudela, P. (2011). Emotional conflict in interpersonal interactions. *Neuroimage, 54*, 1685-91.

Ruz, M., Madrid, E., & Tudela, P. (2012). Interactions between perceived emotions and executive attention in an interpersonal game. *Social Cognitive and Affective Neuroscience*.

Ruz, M., Moser, A., & Webster, K. (2011). Social expectations bias decision-making in uncertain inter-personal situations. *PLoS ONE, 6*, e15762

Sanfey, A. G. (2007). Social decision-making: insights from game theory and neuroscience. *Science, 318*, 598-602.

Scharlemann, J. P. W., Eckel, C. C., Kacelnik, A., & Wilson, R. K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology, 22*, 617-640.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburg: Psychology Software Tools, Inc.

Tottenham, N., Tanaka, J.W., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A., Marcus, D.J., Westerlund, A., Casey, B.J., & Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research, 168*, 242-9.

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current Directions in Psychological Science, 18*, 184–188.

Van Kleef, G.A., De Dreu, C.K.W., & Manstead, A.S.R. (2010). An interpersonal approach to emotions in social decision-making: The emotions as social information model. In M.P. Zanna (Ed.) *Advances in Experimental Social Psychology, 42* (pp. 45-96). Burlington: Academic Press.

Vonk, R. (1998). Effects of cooperative and competitive outcome dependency on attention and impression preferences. *Journal of Experimental Social Psychology, 34,* 265-288.

Wubben, M.J.J., De Cremer, D., &, van Dijk, E. (2009). How emotion communication guides reciprocity: Establishing cooperation through disappointment and anger *Journal of Experimental Social Psychology, 45,* 987-990.

CAPÍTULO 4

RACE, EMOTION AND TRUST: AN ERP STUDY

The content of this chapter has been published as Tortosa, M. I., Lupiáñez, J., & Ruz, M. (2012). Race, Emotion and Trust: an ERP study. *Brain research*.

ABSTRACT

Faces contain certain cues that can be used to infer the intentions of other people and to formulate beliefs about them. The present study explored the extent to which the race of the partners and their emotional facial expressions influenced participants' decision-making in a Trust Game where race and emotional expression had no actual predictive value regarding the partners' reciprocation rate. Behaviourally, participants shared more money with happy than with angry partners. In two separate experiments, electrophysiological results showed an early interaction between race and emotion in the N170 potential and also in the subsequent P200, which suggests inter-dependent processing of those cues in a social context. Overall, our results suggest that racial and emotional cues exert both independent and also interacting effects in the processing of faces in an interpersonal context.

INTRODUCTION

Social interactions play a pivotal role in our lives. Humans have developed efficient abilities that use social features, such as race and emotional facial expressions, to forecast in a direct or indirect way the actions of other people. These could activate stereotypes and may influence the formulation of beliefs that could be used to generate initial trust and to plan subsequent actions toward others. Much research in social cognitive neuroscience has focused on social perception and how automatic biases drive people's impressions of others. In this sense, evaluations are relatively automatic and might occur without conscious monitoring (Cunningham & Zelazo, 2007; Bargh & Williams, 2006). Many of the cues used to trigger evaluative processes and impressions of others come from their faces. Both race and emotional facial expressions have been thoroughly investigated in social cognition. The current study focuses on these two features to investigate whether and how they modulate cooperative behaviour in social interactions.

Racial information easily activates stereotypes and prejudices. Research using implicit measures of prejudice such as the Implicit Association Test (IAT; Greenwald et al., 1998) has shown that positive and negative attitudes can be activated implicitly, which biases social interactions. However, at the same time interactions with minority-group members may also lead to the activation of egalitarian motives, which could inhibit the influence of the stereotypes on judgments and behaviours (see Bargh & Williams, 2006, for a review). For example, Moskowitz et al., (1999) used implicit measures of activation and posited that holding a goal to be egalitarian toward a particular group prevented the activation of stereotypes. Emotional facial expressions, on the other hand, are immediate indicators of behavioural dispositions in people (Darwin, 1872). Since this information is highly relevant to adapt our social behaviour to the tendencies of others, we have developed efficient mechanisms for detecting their emotional states.

Racial and emotional information is mainly extracted from faces, which are processed in fusiform brain regions through fast and efficient brain mechanisms (Haxby et al., 2000; Frith & Frith, 2012). Electrophysiological studies of face processing have shown that the first electrophysiological potentials to be sensitive to these cues are the P1 and N1 deflections. Both coloured and gray-scaled black vs. white faces already differ around 120 ms after face onset (e.g. Ito & Urland, 2003), although these differences are most likely due to low-level physical differences between racial categories (e.g. Tanskanen et al., 2005). On the other hand, faces also elicit rapid brain responses depending on their specific emotional expression. The categorization of facial emotions can occur at latencies as short as 100 ms. For example, Eimer and Holmes (2007) showed that fearful vs. neutral faces generated significant electrophysiological differences in frontal channels as soon as 120 ms after stimulus onset. Other studies have shown that fearful facial expressions elicit a more pronounced negative N100 component than happy and neutral faces as well as enhanced positive P1 amplitudes (e.g. Luo et al., 2010).

The N170, a negative deflection that peaks later in time at bilateral occipito-temporal sites, has larger amplitude for faces than for other stimulus categories. Electrophysiological (Bentin et al., 1996) and fMRI studies of face processing have shown greater activation at right than at left hemisphere sites, as well as more extensive fusiform gyrus activation in the right hemisphere (Puce et al., 1996; Kanwisher et al., 1997). Several reports have studied the sensitivity of the N170 to the structural encoding of faces, facial identification (Bentin et al., 1996; Bentin & Deouell, 2000), and emotional expression (Eimer & Holmes, 2002, 2007; Herrmann et al., 2002). Nevertheless, the relation of this potential to race and emotion is far from settled. Although some authors have reported that it is not sensitive to these cues, others have shown modulations in either or both directions (see Ito & Bartholow, 2009, for a review).

Similarly, another ERP deflection that occurs during the same temporal window as the N170, the anterior face-sensitive vertex positive potential (VPP), has been described as sensitive to facial stimuli (Jeffreys, 1989). A recent study (Wiese, 2012) examined the influence of race on the VPP and reported larger amplitudes for other-race faces. The VPP also appears to be enhanced for emotive (happy and sad) versus neutral expressions (Luo et al., 2010; Jaworska et al., 2012).

In addition, other authors have included participants of two races, black and whites, looking at black and whites faces, and have reported modulations in the fronto-central N200 potential as a function of the race of the participant (Dickter & Bartholow, 2007). The N200 has been specifically associated with deeper processing of faces, and is typically larger to faces of one's own race than to other races (Ito, Thompson, & Cacioppo, 2004; Ito & Urland, 2003). Emotion modulates the amplitude of the N200 in fronto-central channels, which is typically larger to neutral than fearful faces and for happy than angry facial expressions (Eimer & Holmes, 2002; Kubota & Ito, 2007; Ruz et al., 2012).

In a similar time range of processing, some authors have explored how race modulates the P200 at occipito-temporal sites[3] (Stahl et al., 2008, 2010; Wiese, 2012). In these studies, P200 amplitudes were larger for same as compared to other-race faces. Research looking at the modulation by valence of the occipital P200 has reported an enhanced deflection for negative emotional stimuli (Dennis & Chen, 2007). For example, in an affective evaluation task the P200 related to unpleasant stimuli was more positive than for pleasant stimuli (Delplanque et al., 2004). On the other hand, the amplitude of the P200 is reduced following the presentation of angry, compared to neutral facial expressions, (Horley et al., 2001).

Finally, the modulation in the P300 component by race depends mainly on the task and relies on contextual updating of existing content or on a motivated attempt to resolve inconsistent information (Ito & Bartholow, 2009; Nieuwenhuis, et al., 2005). On the emotion realm, the P300 has been linked to motivationally significant events and it typically reflects the arousing content of stimuli (Keil et al., 2002). It is often conceptualized as indexing attention to motivationally relevant stimuli (Kubota & Ito, 2007), with a larger P300 for positive and negative valences compared to neutral stimuli (Keil et al., 2002).

Despite the fact that under normal circumstances race and facial emotion displays are combined to shape the perception of a single entity, a person, research has scarcely investigated the simultaneous perception of these social cues. As an exception, Kubota and Ito (2007) studied the time course of their processing in a task where people made explicit either racial or emotional categorization judgments (the other dimension being task irrelevant). Based on the lack of evidence of interactions between race and emotion in the electrophysiological components analysed, the authors concluded that both cues are processed quickly, independently and in parallel.

Although previous research has considered task and context effects on ethnicity processing (Wiese et al., 2009; Caharel et al., 2011) they have, however, not explored a crucial context in which the race and emotional expression of others are usually processed, that is, social interactions. As mentioned in previous paragraphs, we often evaluate those pieces of information trying to predict the beliefs and proximate behaviour of others. Thus, adding a social context to the task performed by participants in which predicting the behaviour of others is relevant may affect the way their race and emotional

[3] However, other studies have localized this potential as the positivity occurring at fronto-central sites before the N200 (Ito & Urland, 2003, 2005; Kubota & Ito, 2007; Ito & Bartholow, 2009). In this case, the P200 is considered the same as the VPP potential.

expression are processed, and the interactions between them. Our studies emphasize the social interaction with a partner, which might make both emotion and ethnicity implicitly salient.

Therefore, the main goal of the current study was to explore how the brain processes race and facial emotional expression within a classic inter-personal economic setting, by investigating how the electrophysiological correlates were modulated by those social cues. With that purpose we adapted the Trust Game paradigm (Berg, et al., 1995; Camerer, 2003), in which participants had to choose whether to cooperate or not with black and white unknown partners who displayed different emotional states. Importantly, in our studies the two types of cues were incidental to the primary task, in the sense that they did not carry information regarding the partner's cooperation rates. This feature allowed us to examine their automatic effect on trust behaviour, unbiased by strategic factors introduced by the task contingencies. Nevertheless, in spite of being incidental, we expected race and emotional face features to have an effect on the early processing of faces, and affect the cooperation rate of the observers, at least for emotional features (Eckel & Wilson, 2003; Averbeck and Duchaine 2009; Tortosa Strizhko, Capizzi, & Ruz 2013).

We performed two separate experiments, which were almost identical. The second one had the main purpose of replicating the finding in Experiment 1 of a previously unreported interaction between race and emotion in the N170 potential, which was sustained in the P200 (thus minimizing the risk of a Type I error). Participants from both experiments also performed an Implicit Association Test (Greenwald et al., 1998) after the Trust Game, which helped to evaluate whether they had implicit negative biases toward black people.

METHODS AND MATERIALS

Participants

Participants were all white students from the University of Granada, who received course credits in exchange and signed a consent form approved by the local Ethics committee. Twenty-two (one left-handed, 20 female; 20 years old in average) participated in Experiment 1 and 25 (two left-handed, 14 female; 22 years old in average) in Experiment 2.

Experimental Tasks

Trust game task

In the two experiments, participants performed a trust game task with a multi-round design. In this game there are two players. One player, the *trustor*, is given a fixed amount of money, which s/he can keep or invest with a player s/he does not know (the *trustee*, e.g., the face displayed in the computer monitor). If an investment is made, the amount is multiplied and then the trustee supposedly decides how to share the amount with the investor. The trustee can either reciprocate the investment (returning a part of it to the trustor) or fail to reciprocate. In the latter case, the trustor obtains nothing from that exchange. Participants always played as trustors in the game and their 12 game partners (i.e. the 12 faces) were the trustees. These partners were represented by photos and nothing was said about their different races and the emotions they would display.

The experimental task in Experiment 1 and 2 was the same except for the options of response cooperation. In Experiment 1 participants received a fixed amount of 3 Eur in each trial, and then they could decide either to share it all, part of it (1, 2 or 3 Eur), or not to share it with the partner. In each trial of Experiment 2, participants received a fixed amount of 1 Eur and responses were dichotomous: they either cooperated sharing the full amount (1 Eur) or they did not cooperate by not sharing anything with the partner. The decision to keep the money ended the trial. If the participant shared anything, the partner received the triple of the shared amount (the quintuple in Experiment 2). Then the partner could either keep the entire multiplied amount or give half of it back (thus corresponding the trust of the participant). Feedback about this decision was presented on the computer screen, which concluded the trial. Trials were independent from each other; they always started with the same endowment regardless of the gains or losses in the previous rounds or the identity of the partner.

Participants were informed that they were not playing against real people on-line, but they were told that the behaviour of their partners mimicked responses given by real trustors in previous games. They were also told that beneficial outcomes were more likely if both they and their partners were cooperative and shared the money, and were encouraged to maximize their outcomes. Nevertheless, the partners' responses were random and therefore not predictable from the decisions of the participants.

Stimuli and procedure. In the Trust game, participants viewed faces of 12 distinct people: 6 photographs of black faces (3 female) and 6 photographs of white faces (3 female), each portraying angry, happy and neutral facial expressions, resulting in a total of 36 different face stimuli. Images were selected from the NimStim Set of Facial Expressions (Tottenham et al., 2009). This standardized set (NimStim) lacks trustworthiness ratings; however, participants completed a questionnaire after the task and rated the trustworthiness of every face in a 1 to 7 point scale. We performed a t-test comparing white and black faces, which did not show differences between the two[4]. All images were presented upright in frontal view at the centre of the computer screen over a silver background.

Each trial started with a display of a Eur symbol for 200 ms (2.1 x 1.6º visual angle) to indicate the endowment of money to the participant. It was then replaced by a fixation point (+, 0.7 x 0.7º) for 500 ms, which was followed by a picture of the partner for that trial (6.2 x 8.3º) for 1500 ms. During this interval, participants had to give a response with their dominant hand by pressing on the response pad the keys 1, 2 or 3, depending on how much they wanted to share, or the 0 key in case they chose to keep their endowment. Afterwards, another fixation point appeared for 500 ms, and then it was replaced by a symbolic feedback symbol (1.0 x 1.0º) in the centre of the screen that indicated the partner's decision for that trial. In order to minimize saccadic eye movements, three possible symbols displayed in three different colours were used as feedback: a green "o", a navy "#" and a maroon "*". Their meaning was: "You have decided to keep the money. You receive 3 Eur. Your partner receives 0 Eur."; "You have decided to share and your partner has decided to correspond"; "You have decided to share and your partner has decided not to correspond". The association between specific symbols and their meaning was counterbalanced across participants. On trials where participants did not enter their decision on time (1.5 ms), they saw the message "¡tarde!" (late!). At the end of the trial a larger (1.0 x 1.0º) fixation point (+) remained on the screen for a random duration between 2000 and 3000 ms. During this interval participants could blink if they needed so. The task consisted of 5 blocks with 108 trials each (plus 12 practice trials performed at the beginning of the session). Every participant saw each identity with the 3 emotional expressions, for an approximate task duration of 50 minutes.

[4] T-test for Experiment 1: t=-0.34, df=5, p=0.75; Experiment 2: t=1.00, df=5, p=0.36.

Implicit association task (IAT)

A Black-White IAT (Greenwald et al., 1998) was administered to each participant after they performed the Trust Game. Their task was to categorize animals either as pleasant or unpleasant, and faces as white or black. Participants responded by pressing the “e” and “i” keys with their left and right hands. The IAT had two blocks: prejudice-consistent and prejudice-inconsistent (their order of presentation was counterbalanced). In the consistent condition, responses for black (faces) and unpleasant (animals) were associated with the same key-press, while white and pleasant where associated with the other key-press. Left and right key-presses were counterbalanced for black/white responses while the pleasant/unpleasant responses remained constant. In the inconsistent condition the opposite key assignments were used. Incorrect responses were indicated with an “X” in the centre of the screen following the response.

Accuracy and reaction times were recorded and used to compute IAT scores for each participant. The IAT test is based on the hypothesis that people respond faster when stimuli that are similar in valence are assigned to the same response key in comparison to when stimuli differ in valence (Devine et al., 2002). Thus, to obtain these indexes, we subtracted the average RT in the congruent block from the average RT in the incongruent block, standardized according to the standard deviation (see the algorithm proposed by Greenwald et al., 2003; see also Ibañez et al., 2010). An IAT score near zero supposedly reflects a neutral implicit racial association, whereas a positive score indicates more positive evaluative associations for whites than for blacks.

Stimuli and procedure. In the IAT, the stimuli used were neutral facial coloured pictures of four black and four white faces (in each case, two male and two female faces) from the NimStim Set of Facial Expressions (Tottenham et al., 2002) database. These were the same as the ones employed in the Trust Game. The animals were 8 full-colour pictures from the IAT-set, half of which were unpleasant (mean 3.62; SD = 0.36), and the other half pleasant, (mean 7.70; SD = 0.52; in a range of the scale of 1-10).

Participants completed 64 trials divided in two blocks (each containing 32 consistent and 32 inconsistent trials). Before every block, participants performed 32 practice trials (16 consistent and 16 inconsistent) where they could see above and below the picture a label with a reminder of the correct key-press response. The order in which the pictures were presented within each block was randomized for each participant.

Electrophysiological recordings and analysis

Presentation of stimuli was controlled via PC running Biological E-prime software (Schneider et al., 2002) connected to a 17-inch monitor. This computer was connected to a Macintosh, which recorded continuous EEG. Electrophysiological data were collected from AgCl electrodes placed on the scalp using a 128-channel Geodesic Sensor Net, connected to a high-input impedance amplifier (200 MΩ). Eye movements were monitored by horizontal and vertical electro-oculogram (EOG) electrodes lateral to and below both eyes. Impedance was measured for all channels and was maintained below 50 kΩ as recommended for the Electrical Geodesics high-input impedance amplifiers. Gain and zero calibration were performed prior to the start of every recording. EEG was recorded continuously with a sampling frequency of 250 Hz using the vertex channel as the online reference. The amplifier band-pass was set at 0.1-100 Hz.

Topographical analysis

Prior to voltage analysis, we used Cartool software (developed by Denis Brunet: http://sites.google.com/site/fbmlab/) for studying the spatial distribution of brain electrical activity at successive time points across conditions. This technique is informative regarding differences between conditions in terms of likely underlying neurophysiologic sources. Using a spatio-temporal cluster analysis of the ERP normalized group-averaged data, Cartool provides different stable maps (topographies) that reflect time periods of stable electric field configurations and dissociable functional states of the brain, or microstates (Murray et al., 2008).

In our experiment, face stimuli ERPs were summarized by a limited number of scalp potential fields. The choice of the optimal number of topographies that best explained each data set was based on a cross-validation criterion (Pascual-Marqui et al., 1995). The output of this segmentation analysis consisted of a sequence of different scalp–topography configurations or template maps for each segmentation and condition, Black/White and Angry/Happy/Neutral. This procedure served to guide the selection of the optimal temporal windows for the ERP waveform analysis.

Event-Related Potentials (ERP) analysis

Continuous raw data were filtered offline using a 40Hz low-pass filter. The EEG was segmented 200 ms before the face onset and 800 ms after it to obtain face-locked ERPs. After segmentation the EEG was submitted to software algorithms for identification of artefacts. Artefact rejection criteria were defined as 70μV for eye blinks and eye movements' channels, and as voltage exceeding ±80μV at any other electrode. Each segment was then visually inspected to remove remains of ocular or other artefacts. In addition, trials that did not meet the criteria set for behavioural analysis were rejected. ERPs were baseline-corrected 200 ms prior to the presentation of the targets. Data from individual channels that were consistently bad (more than 20% of the trials) for a specific subject were replaced using a spherical interpolation algorithm (Perrin et al., 1989). After rejections, the mean number of trials per experimental condition and participant was 53.9, with a minimum criterion of 25 artefact-free trials per condition and participant. These artefact-free epochs were then averaged separately for each experimental condition and participant, and then re-referenced off-line to the average (Dien, 1998; Tucker et al., 1994).

The average amplitudes for particular ERP components were compared across stimulus conditions over the time windows revealed by the Cartool segmentation process. Face-locked ERPs were explored focusing on the N100-P100, N170, VPP, N200, P200 and P300, given their relevance in the study of the electrophysiological correlates of race and emotion facial processing (Ito & Bartholow, 2009; Kubota & Ito, 2007).

Electrodes where the potentials of interest were maximal were selected, and the mean amplitude of the peak averaged over the selected channels during the established time windows for each experimental condition was computed and entered in the analyses.

Repeated-measures ANOVAs with the factors Race (2: black, white), Emotion (3: angry, happy, neutral), and Hemisphere where relevant (2: left, right) were conducted for the mean amplitude of each potential for the face analysis. In the first place, analyses were conducted separately for Experiments 1 and 2. However, as most of the effects were replicated across them and to facilitate readability, we present combined analyses with Experiment (1, 2) as a between-subject factor, in both the behavioural and electrophysiological sections. Results that are significant in both experiments and in which there is no interaction of Experiment with any other factor are presented collapsed across experiments. In the cases of interaction, on the other hand, results are reported separately for each experiment.

Where relevant, effects were evaluated using a Greenhouse-Geisser (1959) correction, although uncorrected degrees-of-freedom for these contrasts are reported in the text (Jennings, 1987).

RESULTS

Behavioural performance

Omission trials, those trials where participants answered late (1.9% on average) and reactions times shorter than 200 ms (0.38%) were excluded from the analysis. Cooperation Rates were introduced into an ANOVA with Race (black, white) and Emotion (angry, happy, neutral) as within-subjects factors, and Experiment as between-subject factor. The rate of cooperation was larger in Experiment 1 than in Experiment 2, *F*(1,43)=18.21, p=.001. Participants cooperated, i.e., shared at least some of the money, on 78% of the trials in Experiment 1 and on 61% of the trials on Experiment 2. There was a main effect of Emotion on Cooperation Rates, *F*(2,86)=49.17, p<.001. Participants cooperated less for angry (51.7%) than for neutral (75.4%) emotional expressions, *F*(1,43)=53.53, p<.001, and more for happy (82.5%) than for neutral expressions, *F*(1,43)=12.26, p=.001. This effect of Emotion was independent of Experiment (p>.3).

However, there was a significant interaction between Race and Experiment, *F*(1,43)=5.96, p<.05. Whereas in Experiment 1 the race of the partners did not affect cooperation rates (F<1), in Experiment 2 participants cooperated more with black (64.4%) than with white (57.5%) partners, *F*(1,23)=12.11, p<.01.

Results of the IAT task revealed that participants showed an implicit negative bias towards black partners, which was evident in both experiments (main effect of experiment p>.3), with an average score of 0.82. This value was significantly higher than zero (t=6.40, df=44, p<.001), and means that participants took more time to respond to the inconsistent than to the prejudice-consistent condition.

ERP results

Data from one participant from each experiment had to be removed due to excessive artefacts. Data from these participants were also omitted in the behavioral analyses.

N100 and P100

The N100 potential peaked at 115 ms over frontal and midline sites in Experiment 1 and was analysed in a 95-135 ms time window. It peaked at 130 ms over the same sites in Experiment 2 and was analysed in a 110-150 ms time window. The amplitude of the N100 was modulated by the Race of the partner, *F*(1,43)=30.67, p<.001, with a larger N100 for black (-1.01μV) than for white faces (*M*=-0.68μV). There was a significant Race X Hemisphere interaction, *F*(1,43)=5.29, p<.05. The difference in amplitude among races was greater in the right hemisphere (0.45µV) than in the left (0.25µV). The factor Emotion did not reach significance levels, *F*(2,86)=2.54, p=.08 (see Figure 7, and Table 1 and 2 for all the means and standard deviations for the two experiments separately).

The P100 peaked at the same latencies over occipital sites. The amplitude of the P100 was also only modulated by the Race of the partner, *F*(1,43)=49.47, p<.001, with a larger P100 amplitude for black (*M*=1.96µV) than for white faces (*M*=1.16µV).

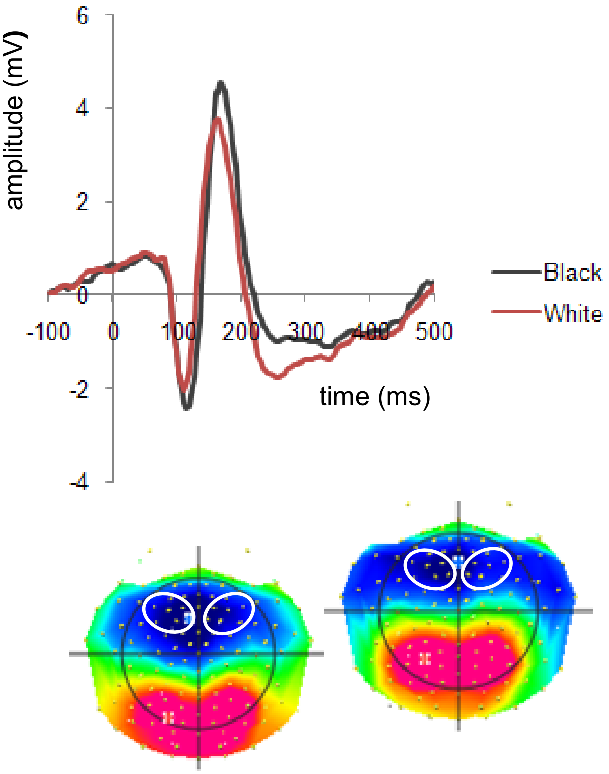


Figure 7. Face-locked ERPs for whites and blacks displaying the N100 and N200 potentials in a frontal channel (F3). The topographies associated to each potential are also displayed, and the white circles highlighting the location of the electrodes used for the amplitude analyses.

N170

The N170 potential peaked at around 170 ms in Experiment 1 and around 200 ms in Experiment 2, at the same temporo-occipital electrode locations. The analysis of the N170 amplitude, in the 150-190 ms time window in Experiment 1 and in the 190-220 ms time window in Experiment 2 revealed a significant Emotion X Race X Hemisphere interaction, *F*(2,86)=3.70, p<.05. Analyses on each hemisphere revealed that the interaction between Emotion and Race was only significant at the right hemisphere, *F*(2,86)=8.19, p<.001 (see Figure 8; at the left hemisphere, p>.2). On the right hemisphere angry and happy expressions showed larger amplitudes (*M*=-2.46µV and -2.63µV, respectively) than neutral expressions (*M*=-2.18µV, *F*(1,43)=9.17 and 18.47, respectively, ps<.01) when displayed by black faces. On the other hand, angry expressions showed larger amplitudes (-2.24µV) than happy (*M*=-1.78µV; *F*(1,43)=16.94, p<.001) and neutral faces (*M*=-1,84µV,*F*(1,43)=7.83, ps<.01) when displayed by whites. There was no difference between happy and neutral faces (F<1)[5]. There was also a main effect of Race, with larger N170s for blacks (-2.43µV) than for whites (*M*=-1.99µV, *F*(1,43)=28.89, p<.001), and an interaction of Emotion x Hemisphere, *F*(2, 86)=6.37, p<.01. While in the left hemisphere there was no effect of emotion, p>.3, in the right hemisphere angry (*M*=-2.35µV, *F*(1,43)=13.55, p<.001) and happy expressions (*M*=-2.21µV, *F*(1,43)=6.09, p=.01) showed larger amplitudes than neutral expressions (-2.01µV). In addition, the Emotion X Experiment interaction was also significant. While the amplitude of the N170 was the same for all emotions in Experiment 1, (*F*<1), it was modulated by this factor in Experiment 2, *F*(2,46)=7.61, p=.001. Planned comparison revealed that angry and happy faces (*M*=-2.32µV and -2.31µV, respectively) elicited larger N170 amplitudes than neutral faces (*M*=-2.02µV; *F*(1,23)=8.66 and 11.13, respectively, ps<.01).

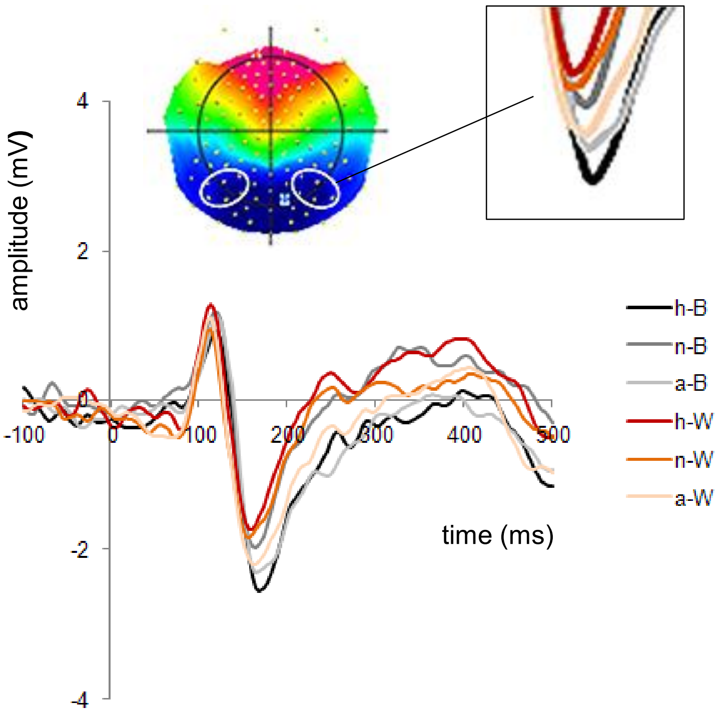


Figure 8. Face-locked ERPs from a posterior–temporal electrode showing the interaction between race and emotion on the N170 potential. The topography during this temporal window is

[5] The interaction of race and emotion at right-hemisphere electrodes was significant in both Experiment 1 and 2 (*F*(2,40)=5.15, p=.01, *F*(2,46)=3.59, p<.05, for Experiment 1 and 2 respectively). Relevant interactions were also all significant in both experiments (all ps<.05).

also displayed, together with white circles highlighting the location of the electrodes used for the amplitude analyses.

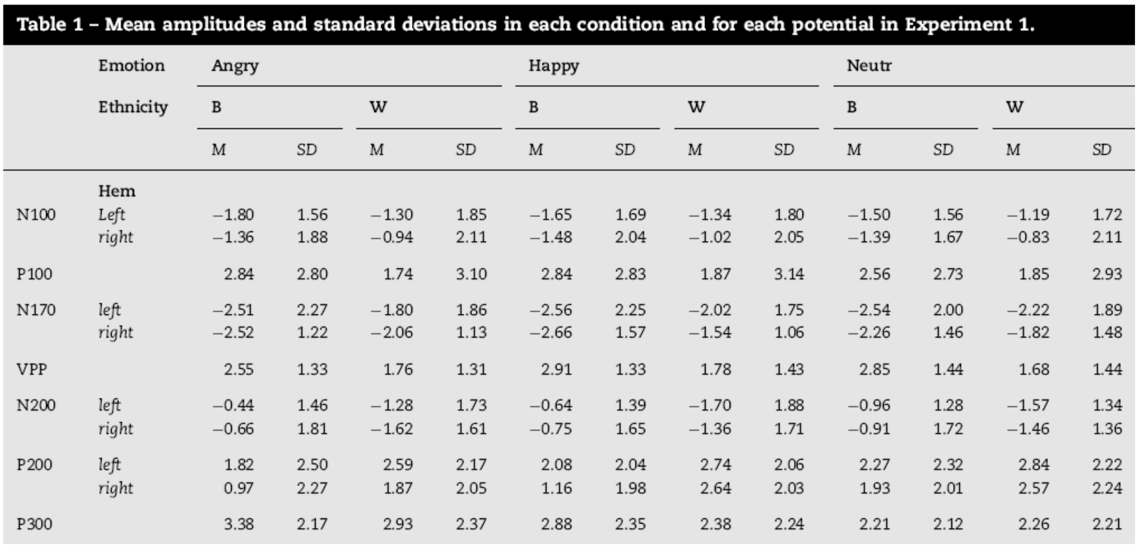


Table 1 – Mean amplitudes and standard deviations in each condition and for each potential in Experiment 1.

	Emotion	Angry				Happy				Neutr			
	Ethnicity	B		W		B		W		B		W	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
	Hem												
N100	*Left*	−1.80	1.56	−1.30	1.85	−1.65	1.69	−1.34	1.80	−1.50	1.56	−1.19	1.72
	right	−1.36	1.88	−0.94	2.11	−1.48	2.04	−1.02	2.05	−1.39	1.67	−0.83	2.11
P100		2.84	2.80	1.74	3.10	2.84	2.83	1.87	3.14	2.56	2.73	1.85	2.93
N170	*left*	−2.51	2.27	−1.80	1.86	−2.56	2.25	−2.02	1.75	−2.54	2.00	−2.22	1.89
	right	−2.52	1.22	−2.06	1.13	−2.66	1.57	−1.54	1.06	−2.26	1.46	−1.82	1.48
VPP		2.55	1.33	1.76	1.31	2.91	1.33	1.78	1.43	2.85	1.44	1.68	1.44
N200	*left*	−0.44	1.46	−1.28	1.73	−0.64	1.39	−1.70	1.88	−0.96	1.28	−1.57	1.34
	right	−0.66	1.81	−1.62	1.61	−0.75	1.65	−1.36	1.71	−0.91	1.72	−1.46	1.36
P200	*left*	1.82	2.50	2.59	2.17	2.08	2.04	2.74	2.06	2.27	2.32	2.84	2.22
	right	0.97	2.27	1.87	2.05	1.16	1.98	2.64	2.03	1.93	2.01	2.57	2.24
P300		3.38	2.17	2.93	2.37	2.88	2.35	2.38	2.24	2.21	2.12	2.26	2.21

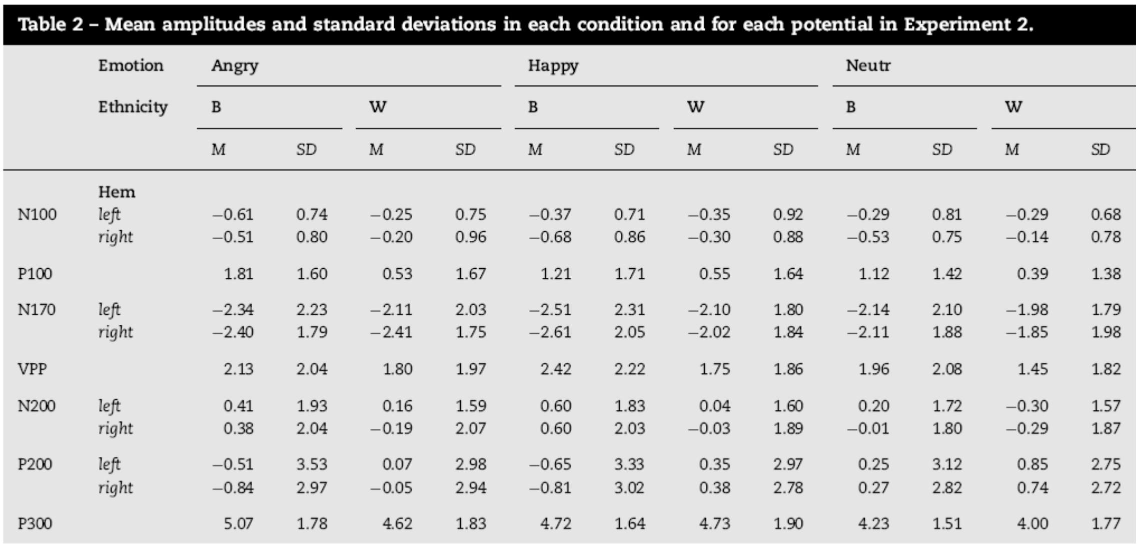


Table 2 – Mean amplitudes and standard deviations in each condition and for each potential in Experiment 2.

	Emotion	Angry				Happy				Neutr			
	Ethnicity	B		W		B		W		B		W	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
	Hem												
N100	*left*	−0.61	0.74	−0.25	0.75	−0.37	0.71	−0.35	0.92	−0.29	0.81	−0.29	0.68
	right	−0.51	0.80	−0.20	0.96	−0.68	0.86	−0.30	0.88	−0.53	0.75	−0.14	0.78
P100		1.81	1.60	0.53	1.67	1.21	1.71	0.55	1.64	1.12	1.42	0.39	1.38
N170	*left*	−2.34	2.23	−2.11	2.03	−2.51	2.31	−2.10	1.80	−2.14	2.10	−1.98	1.79
	right	−2.40	1.79	−2.41	1.75	−2.61	2.05	−2.02	1.84	−2.11	1.88	−1.85	1.98
VPP		2.13	2.04	1.80	1.97	2.42	2.22	1.75	1.86	1.96	2.08	1.45	1.82
N200	*left*	0.41	1.93	0.16	1.59	0.60	1.83	0.04	1.60	0.20	1.72	−0.30	1.57
	right	0.38	2.04	−0.19	2.07	0.60	2.03	−0.03	1.89	−0.01	1.80	−0.29	1.87
P200	*left*	−0.51	3.53	0.07	2.98	−0.65	3.33	0.35	2.97	0.25	3.12	0.85	2.75
	right	−0.84	2.97	−0.05	2.94	−0.81	3.02	0.38	2.78	0.27	2.82	0.74	2.72
P300		5.07	1.78	4.62	1.83	4.72	1.64	4.73	1.90	4.23	1.51	4.00	1.77

VPP

The VPP potential peaked at 170 ms in Experiment 1 and at 200 ms in Experiment 2 at the same centro-medial electrodes. The ANOVA performed on the average voltages across conditions in these electrodes during the 150-190 ms and 190-220 ms time windows, respectively, revealed a significant main effect of Race, *F*(1,43)=56.11, p<.001 due to a larger VPP amplitude for black (*M*=2.47µV) than for white faces (*M*=1.70µV). The interaction Race X Experiment, *F*(1,43)= 6.51, p=01, showed that the difference in amplitude between races was greater in Experiment 1 (1.03µV), than in Experiment 2 (0.51µV). There was also a significant main effect of Emotion, *F*(2,86)=3.45, p<.05. Happy faces (2.21µV) elicited larger VPP amplitudes than neutral ones (*M*=1.98µV;

F(1,43)=7.12, p=.01), whereas angry expressions (*M*=2.06μV) did not differ from happy or neutral faces (ps>.1).

N200

This potential peaked at approximately 248 ms in Experiment 1 and 300 ms in Experiment 2 after target presentation over bilateral frontal sites. It was analysed in the 228-268 ms time window for the former and in the 280-370 ms for the later. There was a main effect of Race, *F*(1,43)=44.19, p<.001, with amplitudes more negative for white (*M*=-0.80μV) than for black faces (*M*=-0.18μV; see Figure 9).

There was also a main effect of Emotion, *F*(2,86)=9.02, p<.001, with amplitudes more negative for neutral (-0.66μV) than for angry and happy expressions (-0.40μV; both ps<.001). There was also a significant interaction between Emotion and Hemisphere, *F*(2,86)=4.27, p=.01. Specifically, N200 amplitudes were larger for neutral (*M*=-0.65μV) than for angry (*M*=-0.28μV, *F*(1,43)=18.61, p<.01) and happy (*M*=-0.42μV, *F*(1,43)=5.25, p<.05) expressions in the left hemisphere and larger for neutral (*M*=-0.67μV) than for happy expressions (*M*=-0.38μV, *F*(1,43)=15.68, p<.001), in the right hemisphere.

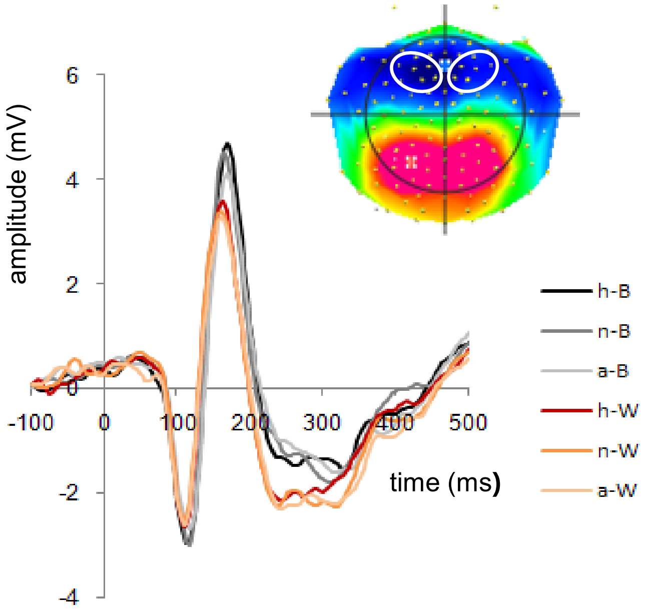


Figure 9. Face-locked ERPs from a frontal (F2) electrode showing the N200 potential. The topography during this temporal window is also displayed, together with white circles highlighting the location of the electrodes used for the amplitude analyses.

P200

This potential peaked over bilateral occipito-temporal sites at approximately 240 ms in Experiment 1 and at 290 ms in Experiment 2. It was analyzed in the 220-270 ms time window in Experiment 1 and in the 260-310 ms in Experiment 2. It showed a main effect of Race *F*(1,43)=27.70, p<.001, with amplitudes more positive for whites (*M*=1.46μV) than for blacks (*M*=0.66μV), and a main effect of Emotion, *F*(2,86)=37.54, p<.001, with a more positive amplitude for neutral (*M*=1.47μV) than for happy (*M*=0.99μV; *F*(1,43)=34.13, p<.001), and for the later than for angry expressions, (*M*=0.74μV; *F*(1,43)=7.90, p<.01). We also obtained a three-way interaction between Emotion, Race and Hemisphere, *F*(2,86)=4.08, p<.05. This interaction showed a pattern similar to the N170. Separate ANOVAS for each hemisphere showed that the interaction Emotion x Race was only significant at the right hemisphere, (p>.3 at the left location). The P200 was larger for neutral (*M*=1.10μV) than for angry (*M*=0.06μV; *F*(1,43)=61.12, p<.001) and happy expressions (*M*=0.18μV; *F*(1,43)=47.83, p<.001) for black partners. On the other hand, neutral and happy expressions (*M*=1.66μV and 1.51μV, respectively) showed larger amplitudes than angry (*M*=0.91μV; *F*(1,43)=17.65 and 20.41, respectively, ps<.001) when displayed by white faces. There was no difference between happy and neutral faces (F<1).

P300

This potential was analysed in the 428-560 ms and in the 490-570 ms time windows over centro-medial electrodes, for Experiment 1 and 2 respectively. There was a main effect of Emotion on its amplitude, *F*(2,86)=29.42, p<.001. Planned comparison revealed that angry faces (*M*=4.00μV) elicited larger P300 amplitudes than happy expressions (*M*=3.68μV; *F*(1,43)=8.18, p<.01), and the latter elicited larger amplitudes than neutral ones (*M*=3.17μV; *F*(1,43)=20.30, p<.001; see Figure 10)[6].

[6] Additionally, we introduced the variable Gender in all behavioural and ERP analyses of the Experiment 2 to test for potential modulation of this factor. There were no significant effects involving this variable (all ps>.1). Nevertheless, given that only 14 female and 10 male participants were examined, this null result should be taken with caution as it may reflect insufficient statistical power.

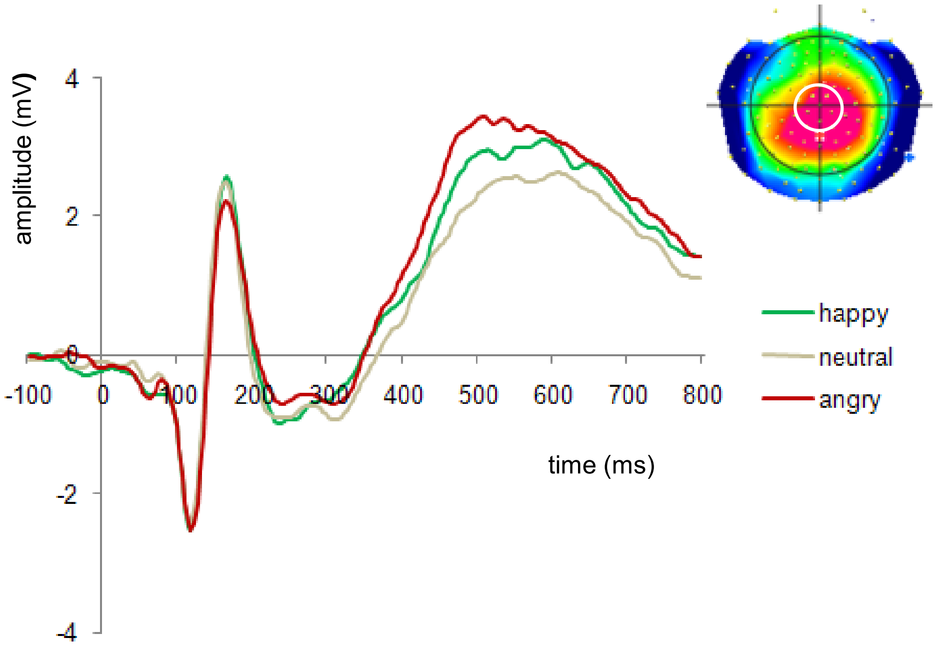


Figure 10. Face-locked ERPs from a central (Cz) electrode showing the effect of emotion on the P300. The topography during this temporal window is also displayed, together with white circles highlighting the location of the electrodes used for the amplitude analyses.

DISCUSSION

The present studies explored how we extract information from other people based on both inherent (race) and variable (emotional expressions) facial characteristics and the extent to which the ERP correlates of face processing in a social decision-making task are modulated by these characteristics. We found previously unreported interactions in the processing of race and emotion on the N170 and P200 potentials, focused on the right hemisphere, which we replicated in a second experiment.

Behavioural data showed higher rates of cooperation with smiling and neutral partners compared to angry ones. The long-lasting effect of emotion even when it was not related to the partner's response of cooperation, replicates previous findings (Scharlemann et al., 2001; Eckel & Wilson, 2003; Tortosa et al., 2013) and points out that facial expressions are strong predictors of others´ behaviour and bias our responses even when our expectations are not matched by evidence (see also Ruz & Tudela, 2011; Ruz et al., 2012).

While race did not affect cooperation rates in the first experiment, the second revealed a bias in favour of black people, and this is so despite the fact that participants showed in both cases an IAT effect indicative of a racial bias that favoured whites. The asymmetry between the effects of race on the two experiments might be related to the change in the responses options, which were dichotomous in Experiment 2 and thus, may have favoured the expression of compensatory behaviour in cooperation rates. This same reason could explain the highest cooperation rate in Experiment 1, where participants had three options of cooperation (1,2,3) vs. one of no-cooperation (0), in comparison with Experiment 2, where the dichotomy in the response option (1-0) made the probability of both choices potentially equal (50%). Nevertheless, the size of the effect of race observed in Experiment 2 is small compared to the effect of facial expression observed in the two experiments, (7% vs. 26% or 36%).

The modulation of racial characteristics of cooperation behaviour in favour of black partners in Experiment 2 could be also explained by the fact that people tend not to show their prejudice-related biases in explicit tasks that manipulate race (Allport, 1954; Monteith et al., 2002), and even tend to compensate it by appearing more "giving" to the exogroup. An alternative explanation could be that the particular black faces selected for the experiment appeared more trustworthy than the white faces used. However, results from the trustworthiness evaluation that participants performed on the faces at the end of the experimental session (see Methods) rule out this explanation, as ratings were equated across categories. Implicit tasks, however (e.g. the IAT, the ''Affective Lexical Priming Score'' or ALPS), have shown evidence of race response-biases even among those who claim to be non-prejudiced (Amodio et al., 2004). IAT scores in our participants showed the existence of such race response-bias in the two samples (Experiment 1 and 2). This positive result in the implicit test might indicate that the way in which our brain initially processes emotional and racial cues escapes conscious control, whereas cooperation responses towards other-race people could be easily controlled to avoid showing non-popular racial biases, as occurs in fact in the cooperation rates participants showed in the trust games.

Turning to the electrophysiological results, an unexpected observation was the delay in the latencies of the peaks of all potentials in the second experiment compared to Experiment 1. This consistent delay was most likely due to uncontrolled differences in background luminance between the two experiments (Wijers et al., 1997). As explained in the Methods section, we tuned the analysis by adapting the temporal windows to the latency of the peaks in each experiment. By using this approach we were able to observe the same pattern of results across experiments, which constitute a replication of our main findings.

In the two studies, ERP results are consistent with past research on the perception of race and emotion, showing initial modulations of ERP components by race and slightly later by emotional expression. These replications validate the utility of our paradigm in measuring the electrophysiological processing of racial and emotional characteristics of facial displays. Race differentially affected early ERP responses on the fronto-central N100, which had larger amplitude for blacks than for whites. This effect may well resemble polarity-reversed effects in the occipital P100, which is known to be sensitive to physical low-level characteristics (Heinze et al., 1994), and thus both potentials may reflect similar processes. We cannot rule out an explanation of our N100 and P100 racial effects in terms of the differences in physical characteristics that naturally occur between the two race categories. Although other studies have reported effects in the same direction as ours even when using grey-scaled stimuli (see Ito & Urland, 2003, 2005) it is relevant to note that using grey scale images do not a priori diminish low-level differences. It has been suggested that, in the context of race, these effects are boosted by early orientation to more novel targets, as a form of rapidly occurring vigilance or greater attention to the racial group (see Ito & Bartholow, 2009; Kubota & Ito, 2007). There is not conclusive research about this issue to date though.

As in the case of the N100, the N170 potential had larger amplitudes for blacks than for whites. This type of modulation could reflect increased demands placed upon the extraction of configural features of other-race faces (Walker et al., 2008). Walker et al. (2008) investigated face-related neural activity and its modulation by race. Participants with greater social contact with the other race showed a diminished difference in N170 mean amplitudes during the structural encoding of whites vs. blacks. It could also reflect an enhanced configural processing for own-race faces (Stahl et al., 2008), as the N170 showed more negative amplitudes for other-race faces. A larger and delayed N170 has also been reported for inverted compared with upright faces (Eimer, 2000; Rossion et al., 2000). Thus, it could also be possible that other-race people might have caused a disruption of configural processing and hence an increase in the amplitude of this potential.

Most importantly, we found a previously unreported interaction in the processing of race and emotion in the N170, which was focused on the right hemisphere and was replicated in a second experiment, with a different sample. The results of our two experiments contradict previous reports of no interactions between these two social cues (Kubota & Ito, 2007), as well as some models of facial processing which propose that invariant and variable face dimensions are processed through parallel and different brain routes (e.g. Haxby et al., 2000). In addition, the subsequent P200 potential displayed the same interactive pattern between racial and emotional facial features. Our results show larger

N170s for both angry and happy expressions than for neutral facial expressions of black partners, whereas white partners generated N170s of larger amplitudes for angry than for happy expressions. Thus, it seems that in our task, the amplitude of this potential was heightened by displays of positive and negative emotions of black people, whereas only the anger of white partners had the same effect. Given that the N170 seems to reflect the categorical processing of faces (Campanella et al., 2000; Aranda, Madrid, Tudela & Ruz, 2010), our results suggest that the faces of black people displaying emotions in a social context, regardless of their valence, may receive heightened processing. This increase in resources would be limited to negative emotions in the case of white partners. The reason for such asymmetry may lie in the differences in familiarity that participants had with people from their own vs. a difference race (Scott & Nelson, 2006). The larger amplitude of the N170 for blacks' angry and happy emotional expressions could also be due to increased attention (Holmes et al., 2003; Olofsson et al., 2008) or deeper processing (Stahl et al., 2008) to emotional in contrast to neutral expressions. According to previous studies showing a modulation of the N170 by facial expression (Batty & Taylor, 2003; Blau et al., 2007), the increased N170 to angry and happy blacks might be caused by the emotional valence of these stimuli (see also Herrmannet al., 2007). It could be the case that participants were more cautious for emotional expressions of black people, regardless of its type. For whites, this enhanced processing would only be linked to partners displaying anger, which is a truly threatening expression. In this case, attention might be oriented towards events that might pose a threat to the perceiver (Öhman & Mineka, 2001).

In addition, and given that the modulation of the N170 appears to be highly task-dependent (Walker et al., 2008; see also Ito & Bartholow, 2009), additional factors for this divergence between the interaction reported here and previous findings could be due to the differences between earlier investigations and the current paradigm. Previous research showed no interaction between emotion and ethnicity of faces, but those results were observed in contexts in which either none of these factors or only one of them were task-relevant (Kubota & Ito, 2007). In contrast, the current task employed a social interaction game in which participants may have used emotional and racial information as relevant sources of information to try to predict the trustworthiness of their partners, and this process may have been different across races. Nevertheless, further research would be needed to clarify the nature of this interesting interaction between emotion and ethnicity in the processing of faces in social contexts.

Similar but attenuated effects were found in the fronto-central VPP, with higher amplitude for black than for whites, in accordance with Wiese´s study (2012), and also for happy emotional expressions. The VPP is a positivity that earlier studies have proposed to show identical response properties to the N170 (Joyce & Rossion, 2005). Although this matter is far from settled (see Ibáñez et al., 2010 for a dissociation between the N170 and VPP), in the current study both potentials indeed reflected the same patterns in the main effects, as well as some of the interactions (Table 3).

The N200 was modulated by both race and emotion, with larger amplitudes for whites than for blacks and also for neutral expressions compared to the other two (but see Ruz et al., 2012). The occipito-temporal P200 potential seems to mirror the effects of the N200. The P200 had larger amplitudes for white than for black partners, its amplitude was also larger for neutral than for angry and happy expressions (see Lucas et al., 2011, for further discussion about the P200 and N200). Those results are not in the same direction as the literature on emotional processing reviewed in the introduction (see Dennis & Chen, 2007; Delplanque et al., 2004); however to date this potential has not been explored in this type of paradigms and effects may depend on several parameters, including the task and the stimuli employed.periments 1 and 2.

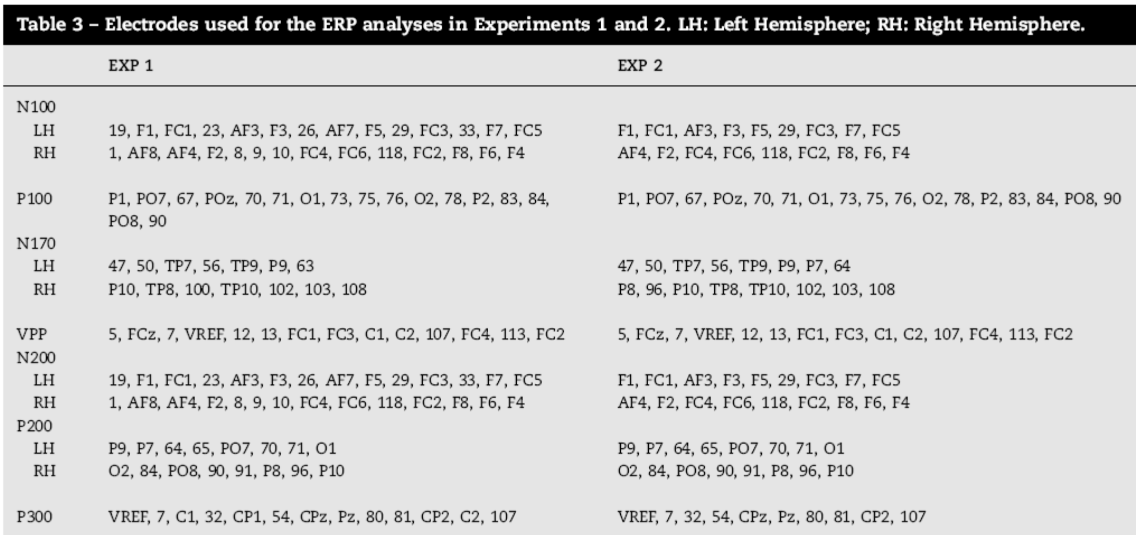


Table 3 – Electrodes used for the ERP analyses in Experiments 1 and 2. LH: Left Hemisphere; RH: Right Hemisphere.

	EXP 1	EXP 2
N100		
LH	19, F1, FC1, 23, AF3, F3, 26, AF7, F5, 29, FC3, 33, F7, FC5	F1, FC1, AF3, F3, F5, 29, FC3, F7, FC5
RH	1, AF8, AF4, F2, 8, 9, 10, FC4, FC6, 118, FC2, F8, F6, F4	AF4, F2, FC4, FC6, 118, FC2, F8, F6, F4
P100	P1, PO7, 67, POz, 70, 71, O1, 73, 75, 76, O2, 78, P2, 83, 84, PO8, 90	P1, PO7, 67, POz, 70, 71, O1, 73, 75, 76, O2, 78, P2, 83, 84, PO8, 90
N170		
LH	47, 50, TP7, 56, TP9, P9, 63	47, 50, TP7, 56, TP9, P9, P7, 64
RH	P10, TP8, 100, TP10, 102, 103, 108	P8, 96, P10, TP8, TP10, 102, 103, 108
VPP	5, FCz, 7, VREF, 12, 13, FC1, FC3, C1, C2, 107, FC4, 113, FC2	5, FCz, 7, VREF, 12, 13, FC1, FC3, C1, C2, 107, FC4, 113, FC2
N200		
LH	19, F1, FC1, 23, AF3, F3, 26, AF7, F5, 29, FC3, 33, F7, FC5	F1, FC1, AF3, F3, F5, 29, FC3, F7, FC5
RH	1, AF8, AF4, F2, 8, 9, 10, FC4, FC6, 118, FC2, F8, F6, F4	AF4, F2, FC4, FC6, 118, FC2, F8, F6, F4
P200		
LH	P9, P7, 64, 65, PO7, 70, 71, O1	P9, P7, 64, 65, PO7, 70, 71, O1
RH	O2, 84, PO8, 90, 91, P8, 96, P10	O2, 84, PO8, 90, 91, P8, 96, P10
P300	VREF, 7, C1, 32, CP1, 54, CPz, Pz, 80, 81, CP2, C2, 107	VREF, 7, 32, 54, CPz, Pz, 80, 81, CP2, 107

The effects of emotion were sustained over the P300, with emotional expressions, both angry and happy, eliciting larger amplitudes than neutral ones. The amplitude of this potential, however, did not vary as a function of the race of the game partners (see Kubota and Ito, 2007). In contrast to early perceptual potentials, the P300 has been mostly linked to decision-making and/or response selection stages of processing (Rugg & Coles, 1995). A number of studies have examined the effects of ethnicity on the P300 and the late positive complex (LPC), which are closely linked (Stahl et al., 2010; Ito &

Urland, 2003, 2005). This potential appears to be sensitive to the context, reflecting context-updating processes that occur in response to explicit attention to task-relevant stimulus features, so that the race effect in those late potentials seems to depend on task and context factors (see also Ruz et al., 2012). Thus, the sensitivity of the P300 to emotion could be related to the influence that this factor exerted on participants' explicit cooperation rates. An additional factor could be the arousing nature of faces displaying emotional expressions, as previous results have shown that P300 amplitudes are enhanced by motivationally significant events (Olofsson et al., 2008), whereas at this late stage of the processing the automatic race bias has been bypassed already.

Overall, electrophysiological results suggest that face processing at early stages is not independent from the emotion and race of partners in a social context. These cues, however, do not exert parallel effects on behavioural responses. Our results highlight the marked effect of emotional facial expressions on cooperation responses, and the inter-wired electrophysiological correlates of race and emotion in early stages of face processing. Further research would be needed to explore the conditions in which the race and emotional expression of faces modulate, in an interactive or independent way, face-related early ERPs components like the N170, and the factors determining whether the race of the partner does or does not modulate overt social behaviours in interpersonal interactions.

REFERENCES

Allport, G.W. (1954). The nature of prejudice. Reading, MA: Addison-Wesley.

Amodio, D.M., Harmon-Jones, E., Devine, P.G., Curtin, J.J., Hartley, S.L., Covert, A.E. (2004). Neural Signals for the Detection of Unintentional Race Bias. *Psychological Science, 15,* 88-93.

Aranda, C., Madrid, E., Tudela, P., Ruz, M. (2010). Category expectations: A differential modulation of the N170 potential for faces and words. *Neuropsychologia, 48,* 4038-45.

Averbeck, B.B., Duchaine, B. (2009). Integration of social and utilitarian factors in decision making. *Emotion, 9,* 599-608.

Bargh, J.A., Williams, E.L. (2006). The Automaticity of Social Life. *Current Directions in Psychological Science, 15,* 1-4.

Batty, M., Taylor, M.J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research, 17,* 613-620.

Bentin, S., Deouell, L.Y. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanisms. *Cognitive Neuropsychology, 17,* 35-54.

Bentin, S., Allison, T., Perez, E., Puce, A., McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience, 8,* 551-65.

Berg, J., Dickhaut, J., McCabe, K. (1995). Trust, reciprocity and social history. Games *and Economic Behavior, 10,* 122-42.

Blau, V.C., Maurer, U., Tottenham, N., McCandliss, B.D. (2007). The face-specific N170 component is modulated by emotional facial expression. *Behavioral and Brain Functions, 3,* 1-13.

Caharel, S., Montalan, B., Fromager, E., Bernard, C., Lalonde, R., Mohamed, R. (2011). Other-race and inversion effects during the structural encoding stage of face processing in a race categorization task: An event-related brain potential study. *International Journal of Psychophysiology, 79,* 266-71.

Camerer, C.F. (2003). Behavioural game theory: Experiments in strategic interaction. Princeton: Princeton University Press.

Campanella, C., Hanoteau, D., Dépy, B., Rossion, R., Bruyer, R., Crommelinck, M., et al. (2000). Right N170 modulation in a face discrimination task: An account for categorical perception of familiar faces. *Psychophysiology, 37,* 796-806.

Cunningham, W.A., Zelazo, P.D. (2007). Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends in Cognitive Sciences, 11,* 97-104.

Darwin, C. (1872). The expression of the emotions in man and animals. London: John Murray.

Delplanque, S., Lavoie, M.E., Hot, P, Silvert, L., Sequeira, H. (2004). Modulation of cognitive processing by emotional valence studied through event-related potentials in humans. *Neuroscience Letters, 356,* 1-4.

Dennis, T.A., Chen, C.C. (2007). Neurophysiological mechanisms in the emotional modulation of attention: the interplay between threat sensitivity and attentional control. *Biological Psychology, 76,* 1-10.

Devine, P.G., Plant, E.A., Amodio, D.M., Harmon-Jones, E., Vance, S.L. (2002). The regulation of explicit and implicit race bias: the role of motivations to respond without prejudice. *Journal of Personality and Social Psychology, 82,* 835-48.

Dickter, C.L., Bartholow, B.D. (2007). Racial ingroup and outgroup attention biases revealed by event-related brain potentials. *Social Cognitive and Affective Neuroscience, 2,* 189-98.

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommentations. *Behavior research methods, instruments & computers, 30,* 34-43.

Eimer, M. (2000). Effects of face inversion on the structural encoding and recognition of faces evidence from event-related brain potentials. *Cognitive Brain Research*, *10,* 145-58.

Eimer, M., Holmes, A. (2002). An ERP study on the time course of emotional face processing. *Neuroreport, 13,* 427-31.

Eimer, M., Holmes, A. (2007).Event-related brain potential correlates of emotional face processing. *Neuropsicología, 45,* 15-31.

Eckel, C.C., Wilson, R.K. (2003). The human face of game theory: trust and reciprocity in sequential games (Vol. 6). New York: Russell Sage Foundation. Greenhouse, S.W., Geisser, S. (1959). On the methods in the analysis of profile data.*Psychometrika, 24,* 95-111.

Frith, C.D., Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology, 63,* 287-313.

Greenwald, A., McGhee, D., Schwartz, J. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74,* 1464-80.

Greenwald, A.G., Nosek, B. A., Banaji, M.R. (2003). Understanding and using the implicit Association Test: I. An improved scoring algorithm*. Journal of Personality and Social Psychology, 85,* 197-216.

Haxby, J.V., Hoffman, E.A., Gobbini, M.I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4,* 223-33.

Heinze, H.J., Mangun, G.R., Burchert, W., Hinrichs, H., Scholtz, M., Münte, T.F., et al. (1994). Combined spatial andtemporal imaging of brain activity during selective attention in humans. *Nature, 372,* 543-6.

Herrmann, M.J., Aranda, D., Ellgring, H., Mueller, T.J., Strik, W.K., Heidrich, A. et al. (2002). Face-specific event-related potential in humans is independent from facial expression. *International Journal of Psychophysiology, 45,* 241-4.

Herrmann, M.J., Schreppel, T., Jager, D., Koehler, S., Ehlis, A.C., Fallgatter, A.J. (2007). The other-race effect for face perception: an event-related potential study. *Journal of Neural Transmission, 114,* 951-7.

Holmes, A., Vuilleumier, P., Eimer, M. (2003). The processing of emotional facial expression is gated by spatial attention: evidence from event-related brain potentials. *Cognitive Brain Research, 16,* 174-84.

Horley, K., Gonsalvez, C., Williams, L., Lazzaro, I., Bahramali, H., Gordon, E., 2001. Event-related potentials to threat related faces in schizophrenia. *International Journal of Neuroscience, 107,* 113-130.

Ibáñez, A., Gleichgerrcht, E., Hurtado, E., González, R., Haye, A., Manes, F.F. (2010). Early Neural Markers of Implicit Attitudes: N170 Modulated by Intergroup and Evaluative Contexts in IAT. *Frontiers in Human Neuroscience, 4,* 1-14.

Ito, T.A., Urland, G.R. (2003). Race and gender on the brain: Electrocortical measures of attention to race and gender of multiply categorizable individuals. *Journal of Personality & Social Psychology, 85,* 616-26.

Ito, T.A., Bartholow, B.D. (2009). The neural correlates of race. *Trends in Cognitive Sciences, 13,* 524-31.

Jaworska, N., Blier, P., Fusee, W., Knott, V. (2012). The temporal electrocortical profile of emotive facial processing in depressed males and females and healthy controls. *Journal of Affective Disorders, 136,* 1072-81.

Jeffreys, D.A. (1989). A face-responsive potential recorded from the human scalp. *Experimental Brain Research, 78,* 193-202.

Jennings, J.R. (1987). Editorial Policy on Analyses of Variance with Repeated Measures. *Psychophysiology, 24,* 474-5.

Joyce, C., Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes: The effect of reference electrode site. *Clinical Neurophysiology, 116,* 2613-31.

Kanwisher, N., McDermott, J., Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception*. Journal of Neuroscience*, *17*, 4302-11*.*

Keppel, G., Zedeck, S. (1989). Data analysis for research designs: Analysis of variance and multiple regression/correlation approaches. Freeman, New York.

Keil, A., Bradley, M.M., Hauk, O., Rockstroh, B., Elbert, T., Lang, P.J. (2002). Large-scale neural correlates of affective picture processing. *Psychophysiology, 39,* 64-49.

Kubota, J.T., Ito, T.A. (2007). Multiple Cues in Social Perception: The Time Course of Processing Race and Facial Expression. *Journal of Experimental Social Psychology, 43,* 738-52.

Lucas, H.D., Chiao, J.Y., Paller, K.A. (2011). Why some faces won´t be remembered: brain potentials illuminate successful encoding for same-race and other-race faces. *Frontiers in Human Neuroscience, 5,* 1-17.

Luo, W., Feng, W., He, W., Wang, N.Y., Luo, Y.J., (2010). Three stages of facial expression processing: ERP study with rapid serial visual presentation. *NeuroImage, 49,* 1857-67.

Monteith, M.M., Ashburn-Nardo, L., Voils, C.I., Czopp, A.M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology, 83,* 1029-50.

Moskowitz, G.B., Gollwitzer, P.M., Wasel, W., Schaal, B. (1999).Preconscious control of stereotype activation through chronic egalitarian goals*. Journal of Personality and Social Psychology, 77,* 167-84.

Murray, M.M., Brunet, M., Michel, C.M. (2008). Topographic ERP Analyses: A Step-by-Step Tutorial Review*. Brain Topography, 20*, 249-64.

Nieuwenhuis, S. et al. (2005). Decision making, the P3, and the locus coeruleus–norepinephrine system. *Psychological Bulletin, 131,* 510-32.

Öhman, A., Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108,* 483-522.

Olofsson, J.K., Nordin, S., Sequeira, H., Polich, J. (2008). Affective picture processing: An integrative review of ERP findings. *Biological Psychology, 77,* 247-65.

Pascual-Marqui, R. D., Michel, C. M., Lehmann, D. (1995). Segmentation of brain electrical activity into microstates: model estimation and validation. *IEEE Transactions on Biomedical Engineering, 42,* 658-65.

Perrin, F., Pernier, J., Bertrand, O., Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology, 72,* 184-7.

Puce,A., Allison,T., Asgari, M., Gore, J.C., McCarthy, G. (1996)*.* Differential sensitivity of human visual cortex to faces, letter strings, and textures: a functional magnetic resonance imaging study. *Journal of Neuroscience, 16,* 5205-15.

Rossion, B. Gauthier, I., Tarr, M.J., Despland, P., Bruyer, R., Linotte, S. et al. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport, 11,* 69-72.

Rugg, M.D., Coles, M.G.H. (1995). Electrophysiology of mind: event-related brain potentials and cognition. Oxford University Press.

Ruz, M., Madrid, E., & Tudela, P. (2012). Interactions between perceived emotions and executive attention in an interpersonal game. *Social Cognitive and Affective Neuroscience*.

Ruz, M., Tudela, P. (2011). Emotional conflict in interpersonal interactions. *Neuroimage, 54*, 1685-91.

Scharlemann, J.P.W., Eckel, C.C., Kacelnik, A., Wilson, R.K. (2001).The value of a smile: Game theory with a human face. *Journal of Economic Psychology, 22*, 617-40.

Schneider, W., Eschman, A., Zuccolotto, A. (2002). E-Prime reference guide. *Psychology Software Tools.*

Scott, L.S., Nelson, C.A. (2006). Featural and configural face processing in adults and infants: a behavioral and electrophysiological investigation. *Perception, 35,* 1107-28.

Stahl, J., Wiese, H., Schweinberger, S.R. (2008). Expertise and own-race bias in face processing: an event-related potential study. *Neuroreport, 19,* 583-7.

Stahl, J., Wiese, H., S.R. (2010). Learning task affects ERP-correlates of the own-race bias, but not recognition memory performance. *Neuropsychologia, 48,* 2027-40.

Tanskanen, T., Nasanen, R., Montez, T., Paallysaho, J., Hari, R. (2005). Face recognition and cortical responses show similar sensitivity to noise spatial frequency. *Cerebral Cortex, 15,* 526-34.

Tortosa, M. I., Strizhko, T., Capizzi, M., Ruz, M. (2013). Interpersonal effects of emotion in a multi-round Trust Game. *Psicológica, 34*, 179-198.

Tottenham, N., Tanaka, J.W., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A. et al. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research, 168,* 242-9.

Tucker, D. M., Liotti, M., Potts, G. F., Russell, G. S., Posner, M. I. (1994). Spatiotemporal analysis of brain electrical fields. *Human Brain Mapping, 1,* 134-52.

Walker, P.M., Silvert, L., Hewstone, M., Nobre, A.C. (2008). Social contact and other-race face processing in the human brain. *Social Cognitive and Affective Neuroscience, 3,* 16-25.

Wiese, H., Stahl, J., Schweinberger, S.R. (2009). Configural processing of other-race faces is delayed but not decreased. *Biological Psychology, 81,* 103-9.

Wiese, H. (2012). The role of age ethnic group in face recognition memory: ERP evidence from a combined own-age and own-race bias study. *Biological Psychology, 89,* 137-47.

Wijers, A.A, Lange, J.J, Mulder, G., Mulder, L.J. (1997). An ERP study of visual spatial attention and letter target detection for isoluminant and nonisoluminant stimuli. *Psychophysiology, 4*, 553-65.

CAPÍTULO 5

THE FRN TRACKS THE EFFECT OF EMOTIONAL DISPLAYS ON COOPERATION IN A TRUST GAME

ABSTRACT

We studied the modulations of facial emotional expressions on the Feedback Related Negativity (FRN) during a Trust Game, in which participants processed outcomes provided by partners displaying different emotions. In two experiments, we observed that non-predictive emotions modulated the amplitude of the FRN potential and also the cooperation rates of participants. In a third experiment, in which the identity of the partners reliably predicted their reciprocation rates, emotional facial expressions still influenced the behavioural cooperation of participants but to a lesser degree. In addition, the absence of an interaction between emotion and the valence of feedback suggests that having a reliable predictor diminishes the generation of explicit expectations from the emotions expressed by other people.

INTRODUCTION

Most times, humans learn from the consequences of their actions (Thorndike, 1911/1970). Our brains predict future events to plan behaviour, monitor performance and detect whether the outcomes of actions meet predictions or are rather better or worse than expected. To achieve that, we rely on flexiblemonitoring mechanisms that make feedback information useful for behavioural adjustments (Holroyd & Coles, 2002; Holroyd, Nieuwenhuis, Yeung, & Cohen, 2003). In support of this, previous research has shown that the brain responds differentially to positive and negative feedback (Nieuwenhuis, Holroyd, Mol, & Coles, 2004; Hajcak, Moser, Holroyd, & Simons, 2006). Part of this neural activity related to errors and negative outcomes is indexed by the Feedback-related negativity (FRN).

The FRN is a negative deflection observed at fronto-central recording sites during electroencephalographic recordings. It peaks between 250-350 ms after feedback onset, and is most pronounced for feedback associated with unfavourable outcomes, such as incorrect responses or monetary loss (Gehring & Willoughby, 2002; Nieuwenhuis, Yeung, Holroyd, Schurger, & Cohen, 2004b). Functionally, it appears to reflect the evaluation of the motivational significance of ongoing events (Gehring & Willoughby, 2002; Masaki, Takeuchi, Gehring, Takasawa, & Yamazaki, 2006; Van Meel & Van Heijningen, 2010). The discrepancy between outcomes and prior predictions is often termed prediction error (Liao, Gramann, Feng, Deák, & Li, 2011). In this context, the FRN would be the reflection of a reward prediction error signal (Holroyd & Coles, 2002) that is used by the brain to make the corresponding adjustments to allow improved efficiency in similar future situations.

Recent studies have shown the crucial role that expectations play on FRN amplitudes (e.g. Bellebaum, Polezzi, & Daum, 2010; Holroyd, Krigolson, Baker, Lee, & Gibson, 2009; Bellebaum & Daum, 2008; Holroyd & Krigolson, 2007). More specifically, some authors have proposed that the FRN is elicited when an error-processing system detects events that are worse than expected (Holroyd & Coles, 2002; Nieuwenhuis et al., 2004a). For example, in Holroyd and Coles' study (2002) unexpected negative feedback was associated with a larger negativity than unexpected positive feedback. In the same line, Hajcak, Moser, Holroyd, and Simons (2007) found that when the feedback signaled loss, FRN amplitudes were larger for trials where subjects expected gains than on those where they actually expected losses. Crucially, Bismark, Hajcak, Whitworth, and Allen (2012), showed that for the FRN to appear in response to negative outcomes, expectations needed to be developed to be then either confirmed or violated. Moreover, the neural system underlying the FRN may monitor event-outcome contingencies even in cases where no objective relationship exists between the outcome and the participants' choice, which would secure the availability of this information for potential future decisions (Bismark et al., 2012).

Few studies have considered the sensitivity of the FRN to social contexts. Van Meel and Van Heijningen (2010) found differential effects in FRN amplitudes as a function of whether people thought they were competing with someone or not in a probabilistic learning task; the FRN was more negative after negative than positive feedback but only when participants were engaged in an interpersonal competition.

In the absence of information about other people, the emotions displayed by someone act as a crucial generator of default inter-personal expectations and signals of social intentions. These clues convey essential information to the receiver about the beliefs and intentions of the sender (Keltner & Haidt, 1999). In this line, several studies (Scharlemann, Eckel, Kacelnik, & Wilson, 2001; Krumhuber, Manstead, Cosker, Marshall, Rosin, & Kappas, 2007) employing economic settings such as the Trust Game (Berg, Dickhaut, & McCabe, 1995) have shown that the emotions displayed by other people bias our social decisions. Initial trust choices are higher when sharing money with happy than with neutral partners (Scharlemann et al., 2001). In addition, happy partners increase cooperative behavior whereas angry partners decrease it, even in game contexts in which emotional displays are no real predictors of the partner's actions (Tortosa, Lupiañez, & Ruz, 2013a; Tortosa, Strizhko, Capizzi, & Ruz, 2013b) or they are incongruent with learned trustworthiness (Campellone & Kring, 2012). Also, when the meaning of emotions in a game conflicts with their natural consequences, error rates increase and responses are slowed down (Ruz & Tudela, 2011; Ruz, Madrid, & Tudela, in press).

However, as far as we know, the FRN has only been seldom investigated in relation to emotional displays in social contexts. Along evolution, and also along each individual's life, we learn that happy emotional expressions often lead to positive consequences, whereas angry emotional displays are associated to negative outcomes. Thus, the perception of a happy or angry emotional face during a Trust Game may lead by default (i.e., without the need to establish any contingency) to the generation of expectations regarding the cooperative (positive) or non-cooperative (negative) future behaviour of a game partner, which could generate the observed bias of non-predictive emotional displays on cooperation rates. However, the neural mechanisms by which such emotional bias takes place in interpersonal decisions remain to be tested. Importantly, these expectations should be differentially reflectedin an electrophysiological marker of the processing of the outcome such as the FRN.

As stated above, in a previous electrophysiological study (see Tortosa et al., 2013a; see Chapter 4) we employed a Trust Game paradigm in which participants had to choose whether to cooperate or not with unknown partners who displayed different emotional states (angry, happy or neutral). In two separate experiments we observed a sustained and prominent effect of the emotional expression over the cooperation rate of the participants, even though emotions were in no manner predictive of the partners' cooperation rates. These effects were also reflected in emotion modulations of early (N170, VPP) and late (P300) event-related potentials (ERP) locked to the partner's face.

In the present study, we present data from the two electrophysiological studies reported by Tortosa et al. (2013a; Experiment 1A and 1B in the current paper) together with a new experiment (Experiment 2). We now focus on the modulations of facial emotional displays on the FRN observed when participants received feedback regarding the cooperative vs. non-cooperative behavior of their partners in the game, which was not explored in the paper by Tortosa et al. (2013a). Importantly, in Experiment 1A and 1B the identity of partners was not predictive of their future rate of reciprocation. That is, the probability of reciprocation of all partners was the same regardless of their identity and of the facial emotion they displayed. Therefore, in these experiments there was no cue (neither identity nor emotion) actually predicting the partners´ cooperative or non-cooperative future behaviour. Nevertheless, as stated above emotions had a marked influence on participant's trust decisions. In experiment 2, in contrast, the identity of the partners became predictive of their cooperation rates. Based on results from pilot studies in our lab, we expected that emotions would no longer influence explicit social cooperation choices, or that their effect would be clearly diminished. We reasoned that the predictive value of the identity would shadow the effect observed previously for emotions displayed by non-predictive partners.

Based on previous literature, in all the experiments we expected a larger FRN for negative (non-reciprocal) than for positive (reciprocal) feedback. And crucially, given our hypothesis that emotions bias decision-making because they generate social expectations of cooperation (happiness) and non-cooperation (anger), we expected that the emotional displays of the game partners would modulate FRN amplitudes in Experiments 1A and 1B, where there was a clear effect of emotion on participant's cooperation rates. However, no modulation of FRN amplitudes was expected to occur in

Experiment 2, as the identity of the partners was a reliable predictor of their cooperation rates.

Experiments 1A and 1B

METHOD

Participants

Participants were all white students from the University of Granada, who received course credits in exchange for their participation and signed a consent form approved by the local Ethics committee. Twenty-two (one left-handed, 20 female; 20 years old in average, SD=4.02) participated in Experiment 1A and twenty-five (two left-handed, 14 female; 22 years old in average, SD=2.25) in Experiment 1B.

Trust Game task

In the two experiments, participants performed a Trust Game task with a multi-round design. In this game there are two players. One player, the *trustor*, is given a fixed amount of money, which s/he can keep or invest with one of twelve different partners (the *trustees)*, which are represented by faces displayed in the computer monitor. If an investment is made, the amount is multiplied and then the trustee decides how to share the amount with the investor. The trustee can either reciprocate the investment (returning a part of it to the trustor) or fail to reciprocate. In the latter case, the trustor obtains nothing from the exchange. Participants always played as trustors in the game and their 12 game partners were the trustees. Nothing was said about the different emotions that partners would display.

The experimental task in Experiments 1A and 1B was the same except for the options of response cooperation. In Experiment 1A participants received a fixed amount of 3 EUR in each trial, and then they could decide either to share it all, part of it (1 or 2 EUR), or not to share it with the partner. In each trial of Experiment 1B, participants received a fixed amount of 1 EUR and responses were dichotomous: they either cooperated sharing the full amount or they did not cooperate by not sharing anything with the partner. The decision to keep the money ended the trial. If the participant shared anything, the partner received the triple of the shared amount (the quintuple in Experiment 1B). Then the partner could either keep the entire multiplied amount or give half of it back (thus corresponding the trust of the participant). Feedback about the participant’s choice and the partner’s decision to reciprocate or not was presented on the computer screen, which concluded the trial. Trials were independent from each other; they always started with the same endowment regardless of the gains or losses in the previous rounds or the identity of the partner.

Participants were informed that they were not playing against real people on-line, but they were told that the behaviour of their partners mimicked responses given by real trustors in previous games. They were also told that beneficial outcomes were more likely if both they and their partners were cooperative and shared the money, and were encouraged to maximize their outcomes. Nevertheless, the partners' responses were random and therefore not predictable from the decisions of the participants.

Stimuli and procedure

In the Trust Game, participants viewed faces of 12 distinct people: 6 photographs of black faces (3 female) and 6 photographs of white faces (3 female), each portraying angry, happy and neutral facial expressions, resulting in a total of 36 different face stimuli. Images were selected from the NimStim Set of Facial Expressions (Tottenham Tanaka, Leon, McCarry, Nurse, Hare et al., 2009). An additional goal of the original experiments (Tortosa et al., 2013a) was to study whether race had a similar effect to emotion in cooperation rates. However, we consistently observed that race had no effect on cooperation rates or it was negligible (see Tortosa et al., 2013a; see also the null results for gender). Thus, in the current paper we focused our analyses on emotion and collapsed across the race and gender of the partners.

Each trial started with a display of a EUR symbol for 200 ms (2.1 x 1.6°) to indicate the endowment of money to the participant. It was then replaced by a fixation point (+, 0.7 x 0.7°) for 500 ms, which was followed by a picture of the partner for that trial (6.2 x 8.3° on average) for 1500 ms. During this interval, participants had to give a response with their dominant hand by pressing on the response pad the keys 1, 2 or 3, depending on how much they wanted to share, or the 0 key in case they chose to keep their endowment. Afterwards, another fixation point appeared for 500 ms, and then it was replaced by a symbolic feedback for 1000 ms (1.0 x 1.0°) in the centre of the screen that indicated the partner’s decision for that trial. In order to minimize saccadic eye movements, three possible symbols displayed in three different colours were used as feedback: a green “o”, a navy “#” and a maroon “*”. Their meaning was: “You have decided to keep the money. You receive 3 EUR. Your partner receives 0 EUR”; “You have decided to share and your partner has decided to correspond”; “You have decided to share and your partner has decided not to correspond”. The association between specific symbols and their meaning was counterbalanced across participants. On trials where participants did not enter their decision on time (1500 ms), they saw the message "¡tarde!" (late!). At the end of the trial a larger (1.0 x 1.0°) fixation point (+) remained on the screen for a random duration between 2000 and 3000 ms. During this interval participants could blink if they needed so. The task consisted of 5 blocks with 108 trials each (plus 12 practice trials performed at the beginning of the session). Every participant saw each identity with the 3 emotional expressions 45 times, for an approximate task duration of 50 min.

Electrophysiological recordings and analysis

Presentation of stimuli was controlled via PC running Biological E-prime software (Schneider, Eschman, & Zuccolotto, 2002) connected to a 17-inch monitor. This computer was connected to a Macintosh, which recorded continuous EEG. Electrophysiological data were collected from AgCl electrodes placed on the scalp using a 128-channel Geodesic Sensor Net, connected to a high-input impedance amplifier (200 MΩ). Eye movements were monitored by horizontal and vertical electro-oculogram (EOG) electrodes lateral to and below both eyes. Impedance was measured for all channels and was maintained below 50 kΩ as recommended for the Electrical Geodesics high-input impedance amplifiers. Gain and zero calibrations were performed prior to the start of every recording. EEG was recorded continuously with a sampling frequency of 250 Hz using the vertex channel as the online reference. The amplifier band-pass was set at 0.1-100 Hz.

Topographical analysis

Prior to voltage analysis, we used Cartool software (developed by Denis Brunet: http://sites.google.com/site/fbmlab/) for studying the spatial distribution of brain electrical activity at successive time points across conditions. This technique is informative regarding differences between conditions in terms of likely underlying neurophysiologic sources. Using a spatio-temporal cluster analysis of the ERP normalized group-averaged data, Cartool provides different maps (topographies) that reflect time periods of stable electric field configurations and dissociable functional states of the brain, or microstates (Murray, Brunet, & Michel, 2008). In our experiment, feedback-related ERPs were summarized by a limited number of scalp potential fields. The choice of the optimal number of topographies that best explained each data set was based on a cross-validation criterion (Pascual-Marqui, Michel, & Lehmann, 1995). The output of this segmentation analysis consisted of a sequence of different scalp-topography configurations or template maps for each condition. This procedure served to guide the selection of the optimal temporal windows for the feedback-locked ERP waveform analysis.

Event-related potentials (ERP) analysis

Continuous raw data were filtered offline using a 40Hz low-pass filter. The EEG was segmented 200 ms before feedback onset and 500 ms after it to obtain feedback-locked ERPs. After segmentation the EEG was submitted to software algorithms for identification of artefacts. Artefact rejection criteria were defined as 70µV for eye blinks and eye movements' channels, and as voltage exceeding ±80µV in any other electrode. In addition, trials that did not meet the criteria set for behavioural analysis (i.e. those where participants answered late, 1.83%, and reactions times shorter than 200 ms, 0.45%), were rejected. ERPs were baseline-corrected 200 ms prior the presentation of the feedback. Data from individual channels that were consistently bad (more than 20% of the trials) for a specific subject were replaced using a spherical interpolation algorithm (Perrin, Pernier, Bertrand, & Echallier, 1989). The minimum criterion of trials per experimental condition and participant was 22. These artefact-free epochs were averaged separately for each experimental condition and participant, and then re-referenced off-line to the average reference (Dien, 1998; Tucker, Liotti, Potts, Russell, & Posner, 1994). After equalling the number or trials for condition to avoid spurious results due to differential number of observations per condition, (a certain number of trials from the conditions with the greater amount were randomly removed to equate the trials with the smaller number), analyses were performed with a mean number of 53trialsperparticipant in experiment 1A and 43 in experiment 1B.

The average amplitudes for the FRN potential were compared across conditions over the relevant time window (300-340ms) revealed by the Cartool segmentation process and following the GAVE waveform inspection. Feedback-locked ERPs were explored over 16 fronto-central electrodes, (Cp2, 80, Pz, 54, CP1, 7, 107, Cz, C2, 113, Fcz, 13, C1, 32, 81, CpZ) in which the FRN was located. The mean amplitude of the peak averaged over the selected channels during the established time window (300-340ms) was computed and entered in the analyses. Only responses where the participants were cooperative could be analyzed, as otherwise the trial ended without information regarding the partner's behaviour. Repeated-measures ANOVAs with the factors Feedback Type (2: reciprocal, non-reciprocal) and Emotion (3: angry, happy, neutral) were conducted for the mean amplitude of the FRN. In the first place, analyses were conducted separately for Experiments 1A and 1B. However, as the effects were replicated across them and to facilitate readability, we present combined analyses with Experiments (1A, 1B) as a between-subject factor, in both the behavioural and electrophysiological sections (see also Tortosa et al., 2013a). Results that are significant in both experiments and in which there is no interaction between Experiment and any other factor are presented collapsed across experiments. In the cases of interaction, on the other hand, results are reported separately for each experiment. Where relevant, effects were evaluated using a Greenhouse-Geisser (1959) correction, although uncorrected degrees-of-freedom for these contrasts are reported in the text (Jennings, 1987).

RESULTS

Behavioural performance

Data from 4 participants from Experiment 1A (one due to excessive artefacts; 3 for not having enough observations) and 9 from Experiment 1B (two due to excessive artefacts; 7 for not having enough observations) had to be removed from the sample. The behavioural results are the same as those published in Tortosa et al., (2013a; except that data from participants who did not reach the criterion of the minimum number of trials for the FRN analyses are not included –only data from 2 participants were excluded in Tortosa et al., 2013a), but we present them here to aid in the interpretation of the FRN results.

Omission trials, those where participants answered late (after 1.500 ms, 1.83% on average) and reactions times shorter than 200 ms (0.45% of all trials) were excluded from the analysis. Averaged cooperation Rates per condition were introduced into an ANOVA with the Emotion (angry, happy, neutral) as within-subjects factor, and Experiment as between-subject factor. The rate of cooperation was larger in Experiment 1A than in Experiment 1B, *F*(1,32)=10.48, *p*<.01. Participants cooperated, i.e., shared at least some of the money, on 81% of the trials in Experiment 1A and on 68% of the trials on Experiment 1B. There was a main effect of Emotion on Cooperation Rates, *F*(2,64)=35.12, *p*<.001. Participants cooperated less for angry (63.1%) than for neutral (79.4%) emotional expressions, *F*(1,32)=41.95, *p*<.001. The difference between happy (81.4%) and neutral was not significant, *p*=0.20. The effect of Emotion was independent of Experiment (*p*>.3).

FRN results

In line with earlier research on the FRN, we examined fronto-central sites in a time window (300-340 ms) centered around the FRN peak at 320 ms, following the GAVE inspection and according to Cartool topographies (see Figures 11A and 11B). This analysis showed a main effect of Feedback Type. The FRN was larger (i.e. less positive) for non-reciprocal feedback (indicating therefore, economic loss; 2.74μV, SD=2.38) compared to reciprocal feedback (4.10μV, SD=2.96), *F*(1,32)=23.35, *p*<.001. There was no main effect of emotion on the FRN amplitude, *F*<1. However, the interaction between Emotion and Feedback Type was significant, *F*(2,64)=12.84, *p*<.001. The difference between reciprocal and non-reciprocal feedback was greater for angry (2.06μV, SD=1.81) than for happy partners (0.84μV, SD=1.75), *F*(1,32)=20.51, *p*<.001, and larger for angry than for neutral (1.20μV, SD=1.94), *F*(1,32)=11.45, *p*<.01. The size of the difference between reciprocal and non-reciprocal feedback for happy and neutral partners was the same, (*p*=.1).

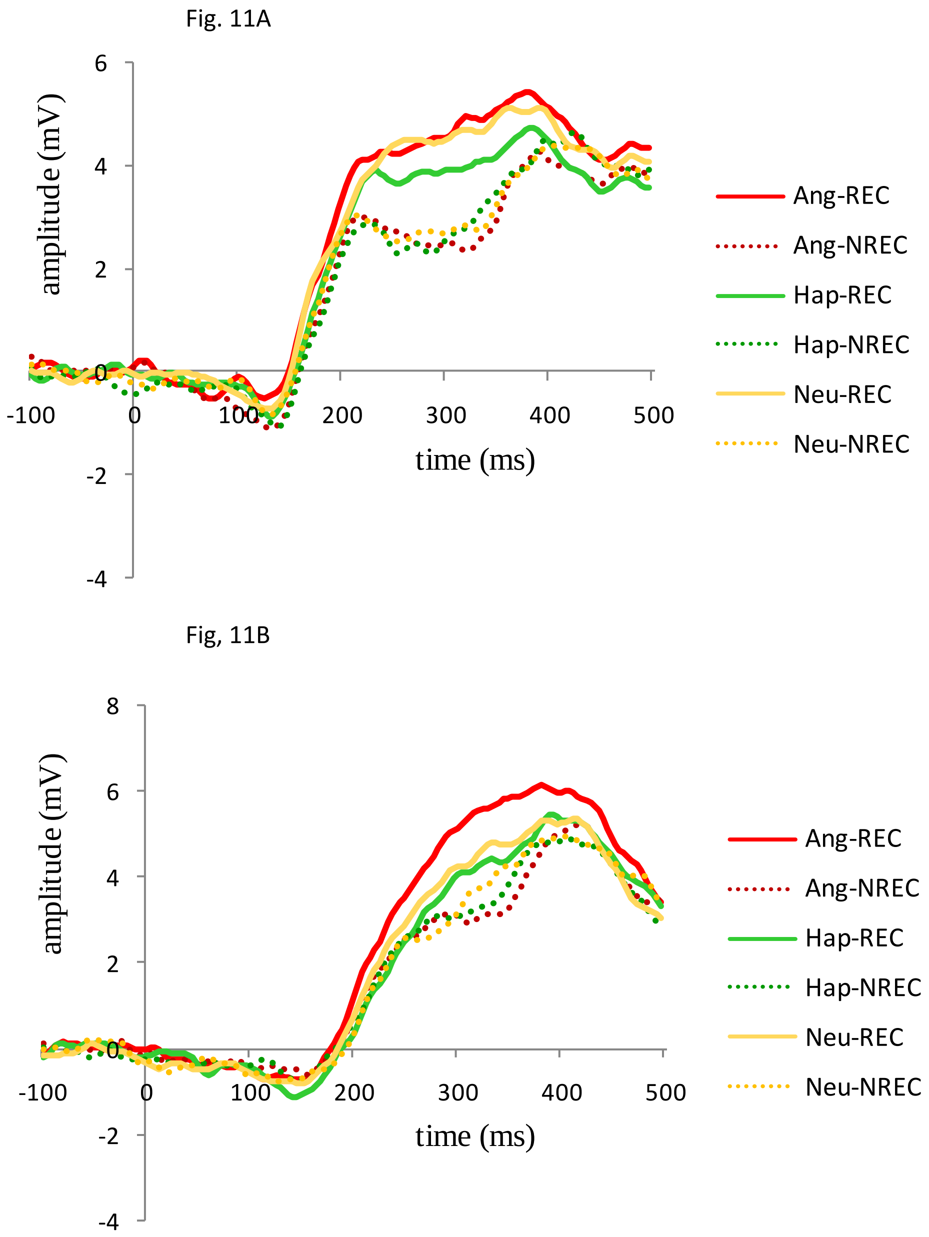


Figure 11A and 11B. Feedback-locked ERP showing the interaction between Emotion (Ang=angry; Hap=happy; Neu=neutral) and Feedback (REC=reciprocal; NREC=non-reciprocal) in the FRN potential (Cz) in Experiments 1A and 1B.

DISCUSSION

In the current experiments we explored the impact of facial emotional expression on the feedback that participants received regarding whether the partners reciprocated their cooperation or not in a Trust Game. Behavioural data showed higher rates of cooperation with smiling and neutral partners, compared to angry ones. The long-lasting effect of emotion even when it was not related to the partner's response of cooperation replicates previous findings (Tortosa et al., 2013b) and points out that facial expressions are strong predictors of others' behaviour and bias our responses even after repeated evidence showing their lack of predictive power (see also Ruz & Tudela, 2011; Ruz et al., 2012).

Turning to the electrophysiological results, we were interested in investigating whether the emotional expression of the game´s partners increased the difference between positive and negative feedback typically observed in the FRN potential. Our results replicated previous studies confirming a larger negative deflection for negative (i.e., non-reciprocal) compared to positive (i.e., reciprocal) feedback (e.g. Nieuwenhuis et al., 2004a; Nieuwenhuis, Slagter, Alting von Geusau, Heslenfeld, & Holroyd, 2005). Importantly, our results extend these effects to the inter-personal social domain. In our experiments, cooperation-related feedback was a positive reciprocation, whereas non-cooperation feedback was a negative reciprocation, as it meant that participants did not "win" any money for that trial.

More importantly, the differential FRN to reciprocal vs. non-reciprocal feedback was greater for angry than for happy or neutral partners. That is, the significance of the outcome (in terms of differences between reciprocal and non-reciprocal responses) was larger for angry partners than for for happy and neutral ones. Given that in the current experiments we could only analyze the FRN after participants' made a cooperative choice, it makes sense that the difference between the two options of reciprocation is heightened after the partner displays an emotional expression that is naturally associated with non-cooperation (i.e. anger, see Ruz & Tudela, 2011; Tortosa et al., 2013b). From a social cognitive point of view, an angry expression would act as an aggressive signal that alerts the individual. Such alerting may regard the feedback as more salient, which would heighten its processing and thus help to learn what is wrong in order to adapt behavior to the situation. This could be reflected in the larger difference between cooperative and non-cooperative feedback delivered by angry partners, compared to happy ones. These results would also be consistent with the idea that the FRN reflects the motivational impact of the outcome event (Gehring & Willoughby, 2002).

The results from this study showed that negative emotions enhance the effect of reciprocation observed in the FRN potential, which suggests that the effect that emotions have in cooperation rates in the Trust Game might be mediated by a differential evaluation of reciprocation outcomes depending on the emotion displayed by the game partners. However, the effect of emotional expression on the FRN might be a non-strategic mere priming of the observation of different emotions unrelated to its effect on cooperation rates. Faces could induce emotions in the player through contagion or empathy (Parkinson, 1996; de Waal, 2008) and this embodying of emotion could then lead to the observed changes in the FRN.

One way to study this connection would be to investigate the effect of emotion on the FRN in conditions under which emotion has no effect on cooperation rates. If the modulation of emotion on the FRN does indeed mediate its effect on cooperation rates, the FRN emotion-related modulation should disappear in a game in which the emotions displayed by the partners had no impact on decisions as well.

To this end, we performed an additional experiment in which the identity of the game partners predicted their cooperative behaviour, while the emotion continued to be non-predictive. In this way, identity acted as a highly predictive cue of probabilistic outcomes (70% vs. 30% reciprocation rates) so that participants had clear *expectations* about the reciprocation tendencies of every partner in the game. Pilot studies in our lab suggested that such manipulation abolishes or diminishes the effect of non-predictive emotions on cooperation rates. In the same line, a recent study (Campellone & Kring, 2012) has shown that under circumstances in which the behaviour of a partner and his emotional display are incongruent (i.e., an angry expression and trustworthy behaviour or a happy expression and untrustworthy behaviour), participants´ decision to trust are guided by the partner reciprocated behavior rather than by their angry or happy expressions.

Consequently, we predicted that the type of person would modulate the cooperation rate of participants (i.e. higher cooperation with cooperative than with non-cooperative partners). In addition, we expected that the highly predictive value of the partners' identity, would shadow the effect of emotions, so that we would not observe an effect of emotion on cooperation rates anymore. Finally, as we considered the modulation of emotion displays on FRN amplitudes as a mediator of the observed behavioural effect, we expected a similar lack of modulation of emotional displays on the FRN potential.

Experiment 2

METHOD

The differences with the previous studies were as follows.

Participants

Twenty-six (one left-handed) white students (19 female; 21 years old in average, SD=6.25) from the University of Granada participated in exchange of course credits. They all signed a consent form approved by the local Ethics Committee.

Trust Game task

Participants played with 8 different trustees (4 female), each portraying either angry or happy facial expressions. We did not present neutral expressions to obtain a higher number of trials per condition. For a single participant, each partner always displayed the same emotion (the association between partner and emotion was counterbalanced across participants). Furthermore, we created expectations of trustworthiness. Four of the partners (two displaying happy and two displaying angry expressions) reciprocated in 70% of the trials, (cooperative type) while the other four did not reciprocate in 70% of the trials (non-cooperative type; the association between partner and trustworthiness was counterbalanced across participants). Hence, the emotional expression was not predictive of the partners´ cooperation rates, as in the previous experiments, but now identity was. Participants were informed of the specific identity-trustworthiness types at the beginning of the experiment, and asked to memorize it.

The feedback stimuli were two symbols (“*”, “#”) displayed in two different colours (blue or purple; counterbalanced across participants). Feedback was provided in every trial, regardless of the decision of the participant. That is, when participants chose to cooperate, the feedback indicated the reciprocation or non-reciprocation of the partner, as in the previous experiments. Alternatively, when participants chose not to cooperate, the feedback indicated through the same symbols whether the partner´s decision would have been to reciprocate or not, had the participants cooperated. Therefore, in contrast to the previous experiments, the feedback was informative of the partner’s trustworthiness on all trials, regardless of the participant’s specific choice.

The timing of the trials was readjusted to increase the total number of trials, by shortening the duration of the fixation points (jittered from 200-600 ms) and the feedback (700 ms). At the end of the trial the interval for blinking (marked with another fixation point) remained on the screen for a random duration between 1750 and 2225 ms. In total, the task had 10 blocks of 80 trials each (800 trial in total), for an approximate duration of 70 minutes. At the beginning of the session, participants received instructions and performed 40 practice trials, where they learned which partners tended to reciprocate and which did not. At the end of the task, participants filled out a Likert-type questionnaire (ranging from 1 to 10) indicating the cooperation rate they perceived for each partner.

After equalling the number or trials per condition to avoid spurious results due to differential number of observations, analyses were performed with amean number of 57trialsperparticipant and condition. For the ERP analysis, repeated-measures ANOVAs with the factors Feedback Type (2: reciprocal, non-reciprocal), Emotion (2: angry, happy) and Partner's type (2: cooperative, non-cooperative) were conducted for the mean amplitude of the peak averaged for the feedback analysis in the same time windows as in Experiments 1A and 1B and at the same fronto-central electrodes.

RESULTS

Behavioural performance

Data from 5 participants had to be removed (3 because either their behavioural rates or the questionnaire indicated that they did not distinguish the cooperative from the non-cooperative partners; 2 for not having enough observations). Omission trials, those trials where participants answered late (after 1.500 ms, 0.40% on average) and trials with RT shorter than 200ms (0.21%) were excluded from the analysis. Cooperation Rates were introduced into an ANOVA with Emotion (angry, happy) and Partner's type (cooperative, non-cooperative) as within-subject factors. Participants cooperated on 51% of the trials. There was a main effect of Partner's type, *F*(1,20)=229.78, *p*<.001, as participants cooperated more with cooperative than with non-cooperative partners (85.6%, SD=0.17 vs. 16.1%, SD=0.21). The effect of Emotion showed a trend that did not reach statistic significance, *F*(1,20)=4.09, *p*=0.06 (49.6% and 52.5% for angry and happy, respectively), as well as the interaction between Type of partner and Emotion *F*(1,20)=3.68, *p*=0.07. Any other effect or interaction was not significant.

Given the non-predicted trend in the effect of emotion on cooperation rates, we performed an additional ANOVA including Experiment (1A & 1B vs. 2) to compare the effect of emotion between experiments. As predicted, there was a significant interaction between Experiment and Emotion, *F*(1,53)=15.62, *p*<.001, indicating that the effect of emotion was larger in Experiments 1A and 1B than in Experiment 2.

FRN results

Given that feedback was informative of the partner's behaviour in all trials (regardless of the participant's choice), we collapsed across participants' responses to study the effect of the trustworthiness expectations (Partner's type) on feedback processing. The ANOVA with the factors Feedback type (reciprocal, non-reciprocal), Emotion (angry, happy), and Partner's type (cooperative, non-cooperative) showed a main effect of feedback, *F*(1,20)=5.10, *p*<.05, as the FRN was more pronounced (less positive) for non-reciprocal (2.35µV, SD=2.45) than for reciprocal feedback(2.87µV, SD=2.44). Also, there was an interaction between Type of partner and Feedback, *F*(1,20)=4.59, *p*<.05. The difference between reciprocal and non-reciprocal feedback was only significant for cooperative partners, *F*(1,20)=9.09, *p*<.01, (3.13µV, SD=2.51 vs. 2.28µV, SD=2.62; see Figure 12). The difference for non-cooperative partners was not significant, F<1. The interaction between Emotion and Type of partner was also significant, *F*(1,20)=6.31, *p*<.05, with more pronounced (less positive) amplitudes for happy than angry non-cooperative partners *F*(1,20)=5.75, *p*<.05, (2.34µV vs. 2.69µV). The difference for cooperative partners was not significant, F<1. Importantly, the interaction between Emotion and Feedback was not significant, (see Figure 13) and neither was the interaction between Emotion, Type of partner and Feedback (all *F*s<1).

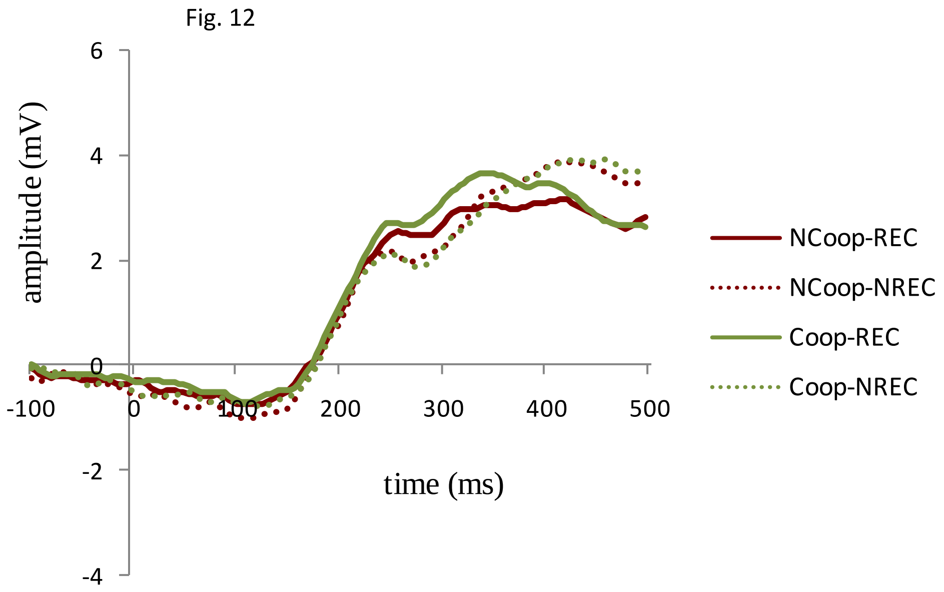


Figure 12. Feedback-locked ERP showing the interaction between Type of partner (NCoop=non-cooperative; Coop=cooperative) and Feedback (REC=reciprocal; NREC=non-reciprocal) in the FRN potential (Cz) in Experiment 2.

Fig. 13

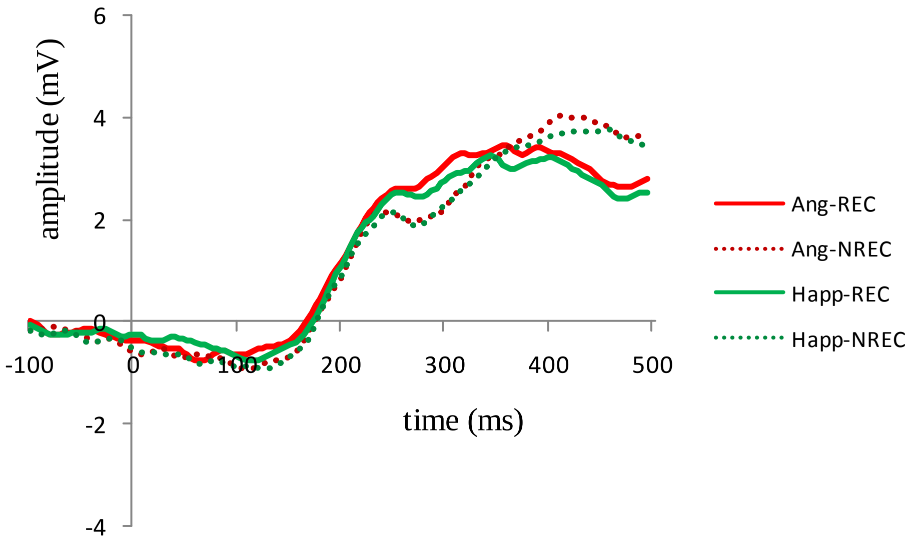


Figure 13. Feedback-locked ERP showing the non-interaction between Emotion (Ang=angry; Hap=happy) and Feedback (REC=reciprocal; NREC=non-reciprocal) in the FRN potential (Cz) in Experiment 2.

DISCUSSION

Results confirmed our hypothesis regarding the lack of effect of emotion on behavioral cooperation rates but only to a certain extent. As we expected, when the identity of the partners was a reliable predictor of his/her cooperative behaviour, trustworthiness played a main role in modulating behavioural responses, with greater cooperation rates with cooperative than with non-cooperative partners. Nevertheless, the effect of non-predictive emotions on cooperation rates remains non-conclusive, as we obtained a trend effect suggesting that emotions still mattered, which was negligible though in comparison with the effect of the type of partner. At the same time, our hypothesis in relation to the lack of modulation of emotion on FRN amplitudes in Experiment 2 was only partially supported. In this case, emotion did not interact with the processing of the feedback. Intriguingly, we observed an unexpected interaction between the type of partner and emotion, indicating that the FRN for non-cooperative partners was more pronounced when they displayed a happy compared to an angry emotion.

For the FRN analysis, we took into consideration both the cooperative and non-cooperative responses of the participants. The FRN was larger again for non-reciprocal (negative) than for reciprocal (positive) feedback, but this was only so with cooperative partners, while with non-cooperative partners there was no difference between feedback types. Note that the current design does not allow disentangling the response of participants (cooperation, non-cooperation) from the type of partner. This is so because our manipulation aimed to generate expectations about people, and participants decided to cooperate or not according to such expectations. In consequence, most of the times participants cooperated with cooperative partners and to chose not to cooperate with non-cooperative partners.

Thus, results showed that the FRN was modulated by the reciprocal/non-reciprocal behaviour only for partners of cooperative type, which were those with which participants chose to cooperate most of the times (see Figure 12). It could be that when people play with non-cooperative partners, they pay less attention to the meaning of the outcome, or they do not generate expectations about what is going to happen next. In any case, the expectations would fit the outcome obtained, as participants´ responses were of non-cooperation most of the times. This would fit with a recent finding proposing that the FRN is not elicited when expectations are not allowed to develop (Bismark, Hajcak, Whitworth, & Allen, 2012), and diminishes when outcomes match expectations, as in this case is expecting a negative outcome and obtaining it (Hajcak et al., 2007). For the non-cooperative partners, an interaction of a different nature arose when they displayed angry vs. happy expressions. This modulation is reflected on the FRN potential, with displayed larger amplitudes for the condition of happy non-cooperative partners. This higher sensitivity of the FRN for both losing and winning, has to do with the nature of the partner rather than to the valence of the feedback. Thus, this affective modulation could emerge as a more intense reaction for happy non-cooperative partners, which in this case is independent of the outcome that they gave. This could be indexing a conflict for untrustworthy partners with happy expressions (see Ruz & Tudela, 2011) as emotions displayed by other people in this case do not predict their natural consequences. On the other hand, as proposed in Averbeck and Duchaine, (2009), our decisions to trust people are based in both social and utilitarian factors, or in this case, the social emotion they express and the actual behavior they manifest. The interaction we observe could be reflecting the association between emotional expressions and the nature of the partner, that together could mediate the processing of social out comes to some extent.

But most importantly, in this second study there was not effect of emotion on positive and negative feedback processing, as indicated by the lack of differences in the FRN for cooperative and non-cooperative responses. This result suggests that having identity as an optimal predictor of the partners’ behaviour, participants were no longer influenced by their incidental emotions when receiving an outcome, and these no longer influenced the outcome-evaluation processes reflected in the FRN.

GENERAL DISCUSSION

In a previous study (see Chapter 4) we explored the extent to which the emotional facial features of other people influenced participants' decision-making in a Trust Game and whether it had an effect on the early processing of faces (see Tortosa el al., 2013a). We have presented here the continuation of that research, in which we took a further step by exploring, through the FRN, how participants processed the feedback received during social interactions and the role that emotional expressions exerted on those correlates.

The uncertainty of experiments 1A and 1B, in which partners displayed different emotions but their identity was non-predictive of their cooperation behavior, led participants to rely on their emotional displays to decide (see Tortosa et al., 2013a; 2013b), and such guidance was reflected in the FRN potential (Experiments 1A and 1B of the current paper). In Experiment 2, in contrast, we expected that emotions would no longer influence behavioral cooperation. Although the marginal effect of emotion in Experiment 2 is not conclusive, the interaction between emotion and experiment shows that the effect of emotion was drastically reduced, if not eliminated. In this experiment the FRN was partially modulated by emotions; but this was independent of the kind of feedback received. This type of interaction could result from a rupture of expectancies of a different nature to the one reported in Experiment 1. The absence of an interaction between emotion and the valence of the feedback, suggests that emotional expressions no longer influence the brain response to feedback information and thus, they are not used to generate explicit expectations.

Our results from E1A and E1B could follow up this argument; in this case, the alerting of the negative feedback from threatening (angry) expressions might had induced a negative affect in the participants and thus the larger FRN under this condition., There is some previous evidence in line with our results. In Santesso and cols´ studies individuals who experienced high punishment sensitivity in response to performance failures, and showed greater negative emotionality had enhanced FRNs for loss feedbacks (Santesso, Dzyundzyak, & Segalowitz, 2011; Santesso, Bogdan, Birk, Goetz, Holmes, & Pizzagalli, 2012). Also in this direction, some studies assessing the affective modulation of the error-related negativity (ERP) component,an ERP time-locked to choice errors that index performance monitoring (Gehring, Goss, Coles, Meyer, & Donchin, 1993), have shown that participants with negative affect displayed larger ERN amplitudes (Luu et al., 2000; Wiswede, Münte, Goschke, & Russeler, 2009a). In this line some studies have examined how induced facial expressions modulate the amplitude of the error related negativity (ERN). Wiswede, Münte, Krämer, and Rüsseler (2009b), reported how the induction of a smile while performing a flanker task led to a reduction of the ERN amplitude. In other studies it was increased after the presentation of negative pictures (Wiswede, Münte, Goschke, & Russeler, 2009a) or under high negative affect and negative emotionality (Luu, Collins, & Tucker, 2000).

The present study has some limitations. First, there are some methodological issues in the design of the tasks that make the comparison between Experiment 1A and 1B difficult, like the option of responses, (degrees of cooperation vs. dichotomic response), and also between Experiment 1 and 2, as only in the latter we could consider both the cooperative and non-cooperative responses of the participants. Second, the marginal effect of emotion in Experiment 2 warrants the need for further experiments with a different sample of participants in which we could study whether emotions still play a role in cooperative decisions even when there are other reliable predictors at hand. Another challenge for future research would be to disentangle the effect of expectations of cooperation vs. non-cooperation of different partners from the actual cooperation vs. no cooperation choices that participant's make, which co-occurred in the present experiment. In addition, future studies could explore how other pieces of information regarding game partners influence implicit and explicit expectations and their effect on feedback processing.

In conclusion, our results show that the FRN tracks the effect of emotional displays, and stress the communicative value of emotion in inter-personal decision-making in conditions where such emotional expressions have different weights for the social interaction. Such influence biases the decision of the participant and this is reflected in the FRN, either by the interaction with the motivational valence of the feedback, or in a further indirect manner, by modulating the expectancies that participants forge according to the identity of the partner.

REFERENCES

Averbeck, B. B., and Duchaine, B. (2009). Integration of social and utilitarian factors in decision making. *Emotion, 9*, 599-608.

Bellebaum, C., and Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *European Journal of Neuroscience, 27*, 1823-1835.

Bellebaum, C., Polezzi, D., and Daum, I. (2010). It is less than you expected: The feedback-related negativity reflects violations of reward magnitude expectations. *Neuropsychologia, 48*, 3343-3350.

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity and social history. *Games and economic behavior, 10*, 122-142.

Bismark, A. W., Hajcak, G., Whitworth, N. M., and Allen, J. J. (2012). The role of outcome expectations in the generation of the feedback-related negativity. *Psychophysiology, 50*, 125-133.

Campellone, T. R., and Kring, A. M. (2012). Who do you trust? The impact of facial emotion and behaviour on decision making. *Cognition and Emotion, 27*, 603-620.

Dien, J. (1998). Issues in the application of the average reference: review, critiques, and recommentations. *Behavior Research Methods, Instruments, & Computers, 30*, 34-43.

Greenhouse, S. W., and Geisser, S. (1959). On the methods in the analysis of profile data. *Psychometrika, 24*, 95-111.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*, 385-390.

Gehring, W. J., and Willoughby, A. R. (2002). The Medial Frontal Cortex and the Rapid Processing of Monetary Gains and Losses. *Science, 295*, 2279-2282

Hajcak, G., Moser, J.S., Holroyd, C.B., and Simons, R.F. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological Psychology, 71*, 148-154.

Hajcak, G., Moser, J. S., Holroyd, C. B., and Simons, R. F. (2007). Its worse than you thought The feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology, 44*, 905-912.

Holroyd, C. B., and Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109*, 679-709.

Holroyd, C. B., and Krigolson, O. E. (2007). Reward prediction error signals associated with a modified time estimation task. *Psychophysiology, 44*, 913-917.

Holroyd, C. B., Krigolson, O. E., Baker, R., Lee, S., and Gibson, J. (2009). When is an error not a prediction error? An electrophysiological investigation. *Cognitive, Affective and Behavioral Neuroscience, 9*, 59-70.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., and Cohen, J. D. (2003). Errors in reward prediction are re£ected in the event-related brain potencial. *Neuroreport, 14,* 2481-2484.

Jennings, J. R. (1987). Editorial policy on analyses of variance with repeated measures. *Psychophysiology, 24*, 474-475.

Keltner, D., and Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition & Emotion, 13*, 505-521.

Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L., and Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion, 7*, 730-735.

Liao, Y., Gramann, K., Feng, W., Deák, G.O., and Li, H. (2011). This ought to be good: Brain activity accompanying positive and negative expectations and outcomes. *Psychophysiology, 48*, 1412-1419.

Luu, P., Collins, P., and Tucker, D. M. (2000). Mood, personality, and selfmonitoring: Negative affect and emotionality in relation to frontal individual differences in the FRN. *Journal of Experimental Psychology: General, 129*, 43-60.

Masaki, H., Takeuchi, S., Gehring, W. J., Takasawa, N., and Yamazaki, K. (2006). Affective-motivational influences on feedback-related ERPs in a gambling task. *Brain Research, 1105*, 110-121.

Murray, M. M., Brunet, M., and Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain topography, 20*, 249-264.

Nieuwenhuis, S., Holroyd, C. B., Mol, N., and Coles, M. G. (2004a). Reinforcement-related brain potentials from medial frontal cortex: Origins and functional significance. *Neuroscience and Biobehavioral Reviews, 28*, 441-448.

Nieuwenhuis, S., Yeung, N., Holroyd, C. B., Schurger, A., & Cohen, J. D. (2004b). Sensitivity of electrophysiological activity from medial frontal cortex to utilitarian and performance feedback. *Cerebral Cortex*, *14*, 741-747.

Nieuwenhuis, S., Slagter, H. A., Alting von Geusau, N. J., Heslenfeld, D. J., and Holroyd, C. B. (2005). Knowing good from bad: differential activation of human cortical areas by positive and negative outcomes. *European Journal of Neuroscience, 21*, 3161-3168.

Pascual-Marqui, R. D., Michel, C. M., and Lehmann, D. (1995). Segmentation of brain electrical activity into microstates: model estimation and validation. *IEEE Transactions on Biomedical Engineering, 42*, 658-665.

Perrin, F., Pernier, J., Bertrand, O., and Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology, 72*, 184-187.

Ruz, M., and Tudela, P. (2011). Emotional conflict in interpersonal interactions. *Neuroimage, 54*, 1685-1691.

Ruz, M., Madrid, E., & Tudela, P. (2012). Interactions between perceived emotions and executive attention in an interpersonal game. *Social Cognitive and Affective Neuroscience*.

Santesso, D. L., Dzyundzyak, A., and Segalowitz, S. J. (2011). Age, sex and individual differences in punishment sensitivity: factors influencing the feedback-related negativity. *Psychophysiology, 48*, 1481-1489.

Santesso, D. L., Bogdan, R., Birk, J. L., Goetz, E. L., Holmes, A. J., and Pizzagalli, D. A. (2012). Neural responses to negative feedback are related to negative emotionality in healthy adults. *SCAN, 7*, 794–803.

Scharlemann, J. P. W., Eckel, C. C, Kacelnik, A., and Wilson, R. K. (2001). The Value of a Smile: Game theory with a human face. *Journal of Economic Psychology, 22*, 617-640.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). E-Prime reference guide. Psychology Software Tools.

Thorndike, E. L. (1911/1970). *Laws and hypotheses for behavior*. In E. L. Thorndike (Ed.), Animal intelligence. Darien, CT: Hafner Publishing Co.

Tortosa, M. I., Lupiañez, J., and Ruz, M. (2013a). Race, emotion and trust: An ERP study. *Brain Research, 1494*, 44-55.

Tortosa, M. I., Strizhko, T., Capizzi, M., and Ruz, M. (2013b). Interpersonal effects of emotion in a multi-round Trust Game. *Psicológica, 34*, 179-198.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., and Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research, 168*, 242-249.

Tucker, D. M., Liotti, M., Potts, G. F., Russell, G. S., and Posner, M. I. (1994). Spatiotemporal analysis of brain electrical fields. *Human Brain Mapping, 1*, 134-152.

Van Meel, C. S., and Van Heijningen, C. C. (2010). The effect of interpersonal competition on monitoring internal and external error feedback. *Psychophysiology, 47*, 213-222

Wiswede, D., Münte, T. F., Goschke, T., and Russeler, J. (2009a). Modulation of the error-related negativity by induction of short-term negative affect. *Neuropsychologia, 47*, 83-90.

Wiswede, D., Münte, T. F., Krämer, U. M., and Rüsseler, J. (2009b). Embodied Emotion Modulates Neural Signature of Performance Monitoring. *Plosone, 4,* e5754.

CAPÍTULO 6

GENERAL DISCUSSION

This doctoral dissertation has explored the impact that the social cues of race and emotional facial expressions of others exert on cooperative decisions during an iterated social exchange setting, the Trust Game. Our goal was to observe the influence of these clues on the decision process and cooperation patterns under different contingencies: 1- in cases where emotional expressions and race of partners in the game were not predictive of their decisions, 2- where there was an association between the emotional expressions of partners and their cooperation rates and 3- where the identity of the partners predicted their trustworthiness, i.e., their reciprocation rates, regardless of their race or emotional expressions.

We have used behavioural and high-density electrophysiological methods (ERPs). Behaviourally, we examined the cooperation rates of participants along each experimental series. At the electrophysiological level, on the one hand, we aimed to study the temporal course of race and emotional expression processing in a social context, focusing in several relevant face-locked ERPs while participants interacted with partners of their same or different race (whites vs. blacks) portraying different facial expressions (angry, happy or neutral). On the other hand, we studied the feedback-related negativity (FRN) in a social context where participants received either favourable or unfavourable feedback regarding the reciprocation of partners in the game, while taking into account the race, emotion and trustworthiness of the partner.

Altogether, we explored some of the cognitive mechanisms that might be at the basis of our social interactions, from a social neuroscience perspective. We further aimed to discern the automatisms from the controlled processes that are engaged when we face others during decision-making.

Overview of the main results of the thesis

As the main behavioural results from this dissertation we observed that the expression of emotions by others is a fundamental clue that people consider when deciding whether to cooperate or not with them. This, however, operates in a flexible way depending on the circumstantial conditions and settings. We will go into details and extend these restrictions along the discussion section.

The main behavioural results in relation to how emotional expressions are implemented in decision-making stress the following points. First, when emotions are not predictive of the behaviour of others and in absence of other reliable information, participants use them to guide their decisions, so that they tend to cooperate more with people portraying happy than angry expressions even after repeated exposure with a lack of contingency between emotions and cooperation patterns of the partners. Second, when the game contingencies conflict with natural expectations (e.g. anger expressions predicting cooperative patterns), learning is slower than when the contingencies fit natural tendencies (e.g. happiness predicting cooperation responses). Third, when the identity of the person becomes a more reliable predictor of cooperation tendencies, participants are still influenced by the emotional expressions of the partner, although the impact of emotions on decisions decreases in a significant manner. However, when participants are explicitly told that emotional expressions are meaningless, and have no social meaning because they are presented randomly by the computer, participants discount their effect and their cooperation rates are unaffected by emotional expressions.

In relation to the race of the partner, our main behavioral results show a consistent automatic or implicit bias pro-whites (as measured by the IAT) that is nevertheless not reflected in the explicit cooperation rates. That is, participants in our experiments tended to cooperate with white and black races alike. Occasionally, there was even a small pro-black bias in cooperation rates, which is consistent with the importance of a controlled process in the expression of explicit attitudes (Lieberman, 2007) This agrees with the finding of a lack of emotional expressions when participants are asked to ignore them. Arguably, to offer a non-racially biased display of themselves, participants might try to avoid their pro-white bias to have an effect on their explicit cooperation rates.

At the electrophysiological level, EEG recordings provided on-line information about the processing of emotional and racial aspects of faces, as well as the processing of the feedback received from the partners in the game. The most important findings concern the interaction between race and emotion in early stages of face processing, which, to our knowledge, had not been previously reported in the literature. We observed this interaction in the N170 and temporo-occipital P200 potentials. Race influenced early ERPs such as the P100 and N100 potentials. Emotion, on the other hand, impacted potentials later in time, around 170 ms after the face presentation, and it continued modulating subsequent electrophysiological deflections such as the P300. Our second main electrophysiological finding relates to the FRN, which was also influenced by the emotional expression of the partner in a consistent way in the cases in which participants employed emotions to generate expectations. However, when the identity of partners became a highly reliable cue predictive of reciprocation of partners, and emotions exerted a much weaker influence over the cooperation rates of the participants, the FRN was only influenced by the emotion displayed by trustworthy partners.

In the following pages we discuss the potential implications of our results in the context of the theoretical models presented in the Introduction section.

The automaticity of facial expression

Are emotions always taken into account when making interpersonal decisions? Our results are not fully clear on this respect. In the experimental series of Chapter 3, results point to a strong association between emotions and social predictions. When facial emotions of the partner are non-contingent with cooperation rates, and in the absence of other predictive cues, participants take them into account to decide whether to cooperate or not with someone. Furthermore, in game versions where there were bias-inconsistent associations between emotions and cooperation patterns (Chapter 3), participants had to exert control over the automatically processed meaning and it took them longer to adapt their responses to the learned associations. The influence of emotional facial displays on the decision-making is quite automatic so far, as it occurs in spite of being truly unpredictive and takes time to control it when made counterpredictive (i.e., when positive emotions predict lack of cooperation).

However, as the second experiment of the series showed, when participant received instructions to ignore the facial emotional displays, then emotions had no effect on explicit cooperation rates. What does this tell us about the automatic effect of emotions? A plausible explanation could be that experimental desirability leads to a vanishing of the effect of emotions. That is, participants do as they are told to do. Therefore, an interim conclusion in this regard might be that the effect that emotions have in our interactions with others is automatic by default, but can be controlled when there is a reason to do so. The behavioural results of our third experimental series, described in Chapter 5, go along these lines. It suggests that when other factors are effective predictors of the trustworthiness of partners, emotions loose part of their automaticity on affecting decisions. However, residual effects seem to persist, so this would show that emotions still play a small role even in the presence of other reliable predictors. This is not surprising as in this setting emotions were not devoid from their social meaning, and they preserved their communicative social value. Emotional expressions thus might be important on-line predictors during current social interactions, within the context of other longer-term predictors such as the identity of the person (when we have cumulative experience with our partner) or group information that we stereotypically attribute to the partner on the basis of their salient features.

In line with this, additional analises on reaction times as well as ongoing experiments on our group suggest this is the case. Being the identity informative about the type of partner, cooperative vs. non-cooperative, and being participants explicitly instructed to ignore their emotional displays, on the basis that they are non-predictive of their cooperative patterns, facial emotional expressions still modulate reaction times. Decision times are longer for incongruent instances of type of partners and emotions (i.e. happy non-cooperative partners and angry cooperative ones) than for congruent associations (i.e. happy cooperative and angry non-cooperative). Hence, irrelevant ignored emotions still influence the time participants need to make decisions in social contexts, creating a conflict that is observed in response times (see Ruz & Tudela, 2011). This could be considered an implicit measurement of the effect of emotions in social contexts.

According to evolutionary theories of emotions, natural selection has shaped emotion programs (Cosmides & Tooby, 2000) in such a way that some emotions are signals interpreted in an automatic manner by others. This is at least the case for basic emotions; those that, on average, produced along evolution a reliable benefit when shared with others (for example, fear was a beneficial signal for survival). Emotional expressions are informative regarding the mental states of others. Thus, the human brain seems to contain highly efficient networks to extract facial expressions of emotions in a fast and highly reliable fashion (Adolphs, 2002) which can be then used to predict the most likely behaviour of others and increase adaptation to the social environment. Our results overall suggest that explicit behaviour, usually guided by facial expressions of emotion and racial bias, can be controlled to discount the effect of these factors when the circumstances require it. However, their incidental processing during social interactions is evident in the neural correlates observed on the ongoing EEG during facial and feedback processing.

Interweaved racial and emotional early processing

One of the most important findings reported in this dissertation concerns the interaction in the processing of race and emotion reflected in early categorical stages of face processing. We have previously discussed in Chapter 4 the possible implications of these interactions, observed in the N170 and P200 ERP potentials, and our goal now is to extend these implications to the theoretical framework described in the introduction of the thesis (Chapter 1).

Classical models of face processing suggest specific systems that contain functional separate routes for the extraction of identity (invariant) and facial expression (variant) at the functional level (Bruce & Young, 1986). Others include the assumption of distinct anatomical routes for variant and invariant aspects, which would be processed independently and with no influence on each other (Haxby, Hoffman & Gobbini, 2002). According to these models, the initial processing of invariant social information, such as race, should not be influenced by variable social information such as emotional expression. Some authors, however, have questioned this assumption concluding that there is no strong evidence for such dissociation (Calder & Young, 2005), and that the point at which these routes bifurcate is not clear. Our electrophysiological results show in fact a differential processing for race and emotional expressions (for example, early effects of race over the P100 and N100 potentials vs. later effects of emotions in the P300), but the fact that at early stages (N170) these cues interact with each other fits better with a view of interactive neural networks (i.e. OFA; Atkinson & Adolphs, 2011), that take part depending of the nature of the stimuli and the task at hand. The mental set that we adopt when facing people to decide whether to trust them or not is qualitatively different than when the task is contextualized outside the social domain, and only requires, for example, a categorization task based on discriminating their facial features (see Kubota & Ito, 2007). Hence, computations engaged during face processing in a social context, such as trust and expectancies, could influence the processes at hand and cause interactions in their neural markers. As no previous research in trust decisions had looked at the joint processing of these social cues, more evidence is needed to corroborate this explanation.

Emotional facial cues and social expectancies in the FRN

One of the goals of the present thesis was to explore how the emotional states of others affect the way we process the feedback related to their behaviour. In the Trust Game, emotions showed by others would lead to the development of expectations in the observer, and these expectations will go in the direction of the natural consequences of the emotion at hand. The reason for that might be that along evolution we have learned that, most of times, emotional expressions are reliable predictors of relevant social consequences.

Along the experimental series of Chapter 5, focused on the FRN, our results showed that this potential is modulated by the emotional expression of the partners. We observed an interaction showing that the FRN was more pronounced when angry partners did not reciprocate than when happy partners delivered the same feedback. Given that the FRN reflects the motivational meaning of a result, we wondered whether a negative feedback delivered from a specific facial expression would augment or decrease the motivational or affective significance of the outcome.

According to the dopaminergic reinforcement learning theory (RL-theory), the magnitude of the FRN decreases in the course of learning (Nieuwenhuis, Holroyd, Mol, & Coles, 2004), insofar as expectancies adjust to actual outcomes. One crucial difference between learning and fully acquired behaviour is the degree of reward unpredictability (Schultz, 1998). Learning happens as long as some outcome is different than predicted, according to classic principles of classical conditioning (Rescorla & Wagner, 1972; Schultz, 2002). At this point, one might wonder which types of feedback were the most expected. We obtained a more pronounced FRN for feedbacks delivered by angry partners, compared to the FRN obtained from happy partners. We could think that the less expected consequence would be a negative (non-cooperative) feedback from a happy person (as happy faces predict favorable consequences), which seems contrary to the RL-theory, which predicts larger amplitudes after the less expected feedback (Nieuwenhuis et al., 2004). However, it is important to note that in the trial analyzed participants always gave an initial cooperative response. Thus, it could be the case that the most expected feedback is a reciprocal response for happy and for angry partners, regardless of their emotion. Even if this interpretation were true, the effect of emotion over the FRN would still remain partially unexplained.

A theoretical line that could bring together the RL-theory and our results relates to the positive or negative affect of the participants while monitoring their execution. The increments on dopamine levels in both the mesencephalic system and the anterior cingulate cortex (ACC) that positive affect generates might counteract the reduction in dopaminergic activity that takes place when an error is committed or a negative feedback occurs (Holroyd & Coles, 2002). In order to investigate this issue Wiswede, Münte, Krämer, & Rüsseler, (2009) induced smiling facial expressions while participant performed a choice reaction task, and measured the ERN produced by errors. The ERN is a negative deflection at frontocentral sites that index performance monitoring and it have been considered to reflect the same neural processes as the FRN (Nieuwenhuis et al., 2004; Miltner Baum, & Coles, 1997; Gehring & Willoughby, 2002). When a positive affect was induced, ERN showed a decrement in amplitude while errors were committed. This is explained by reference to a mechanism of embodying emotions that leads to a change in the error processing.

This supports our argument exposed in the discussion of Chapter 4, referent to the deeper processing when signs of danger, like an angry expression compared to a happy one. Whereas the first one would be amplified, the latter could yield to positive affect and in consequence decrease the amplitude of the FRN. This, together with the lack of a clear expectation of a negative feedback when somebody was angry given that participants had just provided a cooperative decision, could lead to a feasible explanation of the modulations that emotion exerted in the FRN amplitude in Experiments 1A and 1B from the series reported in Chapter 5. In Experiment 2 of the same series, a different interaction arose though, showing that it was the type of partner that modulated the FRN. This could indicate that the clue that participants used to generate expectations has changed, given that identity was now predictive of cooperation rates. In this scenario, emotional expressions adopted a second place.

The main conclusion from the studies presented in the Chapter 5 series emphasizes the ubiquitous quality of emotion in inter-personal decision-making. Our results overall showed additional evidence about the ballistic effect that facial emotions have in social decisions, and supports the automatic nature of emotional processing during interpersonal interactions as shown by an effect over the processing of the interpersonal feedback received. When the identity of the partners has predictive value about their reciprocity, feedback becomes modulated by person identity, but emotional processing is still observed in the FRN. While most of research to date has focused on the intrapersonal effects of emotion, or moods and emotional states of the decider, this dissertation adds to the current literature by studying how the interpersonal effects of emotions affect the FRN, and how these effects are attenuated when others social factors become part of the variables to consider.

Concluding considerations and future research directions.

Many times we make decisions that involve other people without information about what the behaviour of the other person will be. We need signals that allow us to infer their mental states with the purpose of predicting their most likely behaviour and therefore guide us to either an approach or avoidance response. In absence of previous direct knowledge about someone, the signals we seek from him/her include our knowledge about this person' social categories (its gender, age, raze, etc.) and from the transient and immediate cues that his/her face provides. We access the first kind of knowledge, extracted from invariant facial cues, through the stereotypes and social categories we form to manage the world. The members of those categories share some attributes that help to organize, simplify and systematize the complex information in environment we encounter in our daily lives, indispensable for our relation with the outside world. To the other type of knowledge above mentioned, extracted from variant facial cues, we access through theory of mind, essential for our relation with other human individuals, who also have a mental life. Also, most of the times we face with the consequences of our interpersonal decisions and we have to cope with the decisions of others. People use feedback as a self-regulating mechanism to learn and make better decisions in the future.

In the present dissertation we explored and attempted to provide evidence regarding the influence of emotional expressions and race over decision-making within an interpersonal setting as the Trust Game employing behavioural and electrophysiological indices. The results along the 6 experiments reported in the three experimental series of this thesis provide evidence that emotional expressions cause automatic effects over the decision to trust or not someone, and Chapter 3 tells us about the malleability of such effects. The experiments reported in Chapter 4 show that cues like race and emotion can interact at early stages of neural processing. Moreover, the results from Chapter 5 suggest that the FRN is modulated by the emotion of the person we have received a feedback from; besides, these results also show how these effects diminish when a more important factor becomes predictive of the action on course. This can be interpreted as a sign of the malleability of our person-perception abilities, and the underlying flexible system that continually updates our knowledge about others, and allocates different weights to the information that becomes more important depending of the context and the current expectations.

Nevertheless, several research questions are not fully addressed in this work, and they would require further research. We will mention some of the limitations of the research presented in this dissertation and we will finish with proposals of future research that should provide new insights into relevant processes in social decision-making.

First, a main issue remains unsolved. We fail to observe clear effects of the race of the partner in cooperation rates whereas implicit association test (IAT) indices seem to indicate that participants were prejudiced against black people. This could be a manifestation of the dissociation that sometimes emerges between explicit and implicit attitudes (Dovidio et al 1997). Our tasks were designed to explore explicit overt cooperation behaviours rather than implicit biases, and as mentioned before social desirability may have obscured the effect of biases on open cooperation responses. Further research could adapt the Trust Game paradigm to measure implicit biases, for example by having trustworthy and untrustworthy partners of different races and measuring the speed of the decisions in congruent and incongruent conditions, as we are currently doing with emotional expressions.

In this sense, the lack of correlations between IAT indices and cooperation rates in our task, might be due to the equal evaluations in trustworthiness that our participants gave to black and white partners. As other studies have found, individual differences in implicit race attitude correlated with race disparity in trustworthiness evaluations (Stanley, Sokol-Hessner, Banaji, & Phelps, 2011). Following the same argument, Kubota, Li, Bar-David, Banaji, and Phelps, (2013) found that participants accepted more and lower offers from white than from black proposers, in an Ultimatum Game, and that this pattern was accentuated for participants with higher implicit race bias. Their contention is that black American males are stereotyped as untrustworthy (Dotsch, Wigboldus, Langner, & van Knippenberg, 2008) and so differential levels of group-based trust influence economic decision making. Maybe because we did not found lower levels of trust among our participants toward black partners, we fail to find an effect on the allocation of cooperation between groups. Further studies could compare the effect of racial bias obtained with the Trust Game with those observed with other paradigms as the Ultimatum Game, with which an effect of racial bias as been observed (Kubota et al., 2013).

Second, we could seek further specifics about the circumstances in which the emotions of others no longer influence our behaviour and the way in which this is reflected at the neural level.

Third, the design of our second experiment in the last experimental series did not allow disentangling the effect that the trustworthiness of the partner and the cooperation responses may have on the FRN, because participant's responses overlapped with the trust category of the partners (i.e. most responses to trustworthy partners were of cooperative type whereas responses to untrustworthy partners were mostly non-cooperative). Therefore, in the future we should devise a task with a design that allow to dissociate these factors and further clarify our results.

Also, future research could explore the effect of emotions over the FRN when contingencies between emotional expressions and trustworthiness rates are assigned in an intuitive or counter-intuitive manner, as we did in the behavioral series in Chapter 3. An additional interest would be to explore other emotional states such as sadness. Would cooperation rates be modulated by this emotion? Which influence would exert over the FRN and the face-locked potentials?

Social and affective neuroscience is a vast research area that incorporates new and more sophisticated developments at a fast rate. Novel analyses techniques to approach the brain together with advances in integrative models in social cognition and affective-motivational processing, will hopefully provide a better understanding of the cues that affect human social decision-making, the underlying neural systems, relevant computations and the emergence of coherent and adapted behaviour.

REFERENCES

Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and cognitive neuroscience reviews*, *1*, 21-62.

Atkinson, A. P., & Adolphs, R. (2011). The neuropsychology of face perception: beyond simple dissociations and functional selectivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 1726-1738.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology*, *77*, 305-327.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*, 641-651.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. *Handbook of emotions*, *2*, 91-115.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of experimental social psychology*, *33*, 510-540.

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, *295*, 2279-2282.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological psychiatry*, *51*, 59-67.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological review*, *109*, 679-709.

Kubota, J. T., & Ito, T. A. (2007). Multiple cues in social perception: the time course of processing race and facial expression. *Journal of Experimental Social Psychology*, *43*, 738-752.

Kubota, J. T., Li, J., Bar-David, E., Banaji, M. R., & Phelps, E. A. (2013). The Price of Racial Bias Intergroup Negotiations in the Ultimatum Game. *Psychological science*, 0956797613496435.

Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annu. Rev. Psychol.*, *58*, 259-289.

Miltner, W.H., Baum, C.H., & Coles, M.G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a generic neural system for error detection. *Journal of.Cognitive Neuroscience, 9*, 788-798.

Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience & Biobehavioral Reviews*, *28*, 441-448.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, 64-99.

Ruz, M., & Tudela, P. (2011). Emotional conflict in interpersonal interactions. *Neuroimage, 54*, 1685-91.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of neurophysiology*, *80*, 1-27.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, *36*, 241-263.

Stanley, D. A., Sokol-Hessner, P., Banaji, M. R., & Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the National Academy of Sciences*, *108*, 7710-7715.

Wiswede, D., Münte, T. F., Krämer, U. M., & Rüsseler, J. (2009). Embodied Emotion Modulates Neural Signature of Performance Monitoring. *Plosone, 4,* e5754.