Universidad del Desarrollo
Facultad de Gobierno

HUMAN RECIPROCITY

Behavioural components and neurocognitive signatures discriminating outcomes and intentions in trust-repayment behaviour — and some evolutionary considerations

POR: MAURICIO ALEJANDRO ASPÉ SÁNCHEZ

Tesis presentada a la Facultad de Gobierno de la Universidad del Desarrollo para optar al grado académico de Doctor en Ciencias de la Complejidad Social

PROFESORES GUÍA:

Sr. PABLO BILLEKE BOBADILLA

Sr. CARLOS RODRÍGUEZ SICKERT

Julio 2020

SANTIAGO

A Mario.

Table of Contents

List of Figures

1 Prolegomenon (or how the pieces fit)

The purpose of this work is to understand sociality – in all its complexity – from an interdisciplinary perspective, and to use this understanding to shed light on one of its manifestations: human economic behaviour. This is an unaccomplishable task; thus, I resigned to put together five of the six academic articles that I have developed by working (and having a great time) in my Ph.D. studies in Social Complexity Sciences, at the Centro Interdisciplinario de Complejidad Social of Universidad del Desarrollo.

These five articles fit together into a coherent narrative. The first one is a kind of "history of sociality", and could be considered as the general framework of this report. It reviews the evolution of the neurocognitive mechanisms that allows phenomena that range from the formation of groups of bonded conspecifics to the social behaviour of *Homo sapiens* – embedded in complex networks of indirect reciprocity that exists inside social structures with emergent properties. Its central thesis (Figure 1) is that, accomplished some physiological requirements that enable the formation of aggregates of individuals, the prosocial neuropeptides oxytocin and arginine-vasopressin inhibit the activity of the amygdala (a brain structure related with aversive memories), which facilitates the approach to conspecifics, mainly by diminishing social aversion and sensibility to social betrayal, and increasing trust. Upon these social bonds, neural networks related to allostasis are co-opted, via a pluralism of evolutionary forces, by mechanisms of social cognition. The association cortices that emerge in this context – remarkably the prefrontal cortex and the temporoparietal junction – implement processes of cognitive control and mentalisation. This allows the manifestation of the complex structures of cooperation via mechanisms of direct

and indirect reciprocity that we observe in contemporary human societies.

My second work deals specifically with oxytocin and arginine-vasopressin, the prosocial neuropeptides that bind conspecifics. Oxytocin and arginine-vasopressin are two phylogenetically conserved neuropeptides that have been implicated by a large body of research – ranging from rodents to humans – in affiliative and trust behaviors. Thus, this review summarizes the current findings on the associations between polymorphisms in the oxytocin and vasopressin receptors and human sociality, highlighting the genetic contributions of receptor polymorphisms to diseases related to social deficits, in addition to their interactions with genes involved in other pathways (like those of dopamine, serotonin, and reelin). As stated above, the main effect of the signalling pathways of oxytocin and vasopressin is exerted on the volume, connectivity, and activation of specific neural structures – notably, the amygdala, thus reducing the aversion to social betrayal and facilitating social binding. Upon these bonds, neural networks arising under a pluralism of evolutionary forces enable the neurocognitive processes of cognitive control and mentalisation. These processes are required for the manifestation of two important other-regarding preferences: altruism (a costly action that benefits others) and reciprocity (the repayment of acts in kind). Both are among the building-blocks of complex sociality.

While altruism expresses preferences about the outcome of a social interaction, reciprocity requires, in addition, ascribing intentions to others. My third work compares individuals behaviour and neurophysiological activity under outcomes- (i.e, altruistic) versus intention- (i.e., reciprocal) based interactions under different endowments, in the same subject and during the same session. For this, we used a mixed version of the Dictator and the Investment games, together with electroencephalography. The study shows that subjects displayed positive and negative reciprocity depending on

the amount of trust they received. Furthermore, a subject's late frontal negativity differed between conditions, predicting responses to trust in intentions-based trials. Noteworthy, the cortical sources of this activity are structures related with mentalising and cognitive control, namely the temporoparietal junction and the dorsolateral and dorsomedial prefrontal cortices.

As the mentalising and cognitive control systems develop, different other-regarding preferences could arise by the different interaction of both systems during ontogeny. Thus, in my fourth work we seek to extend the conclusions of my third one, by measuring altruism and reciprocity, in addition to cognitive control capabilities, in subjects of different ages, with the objective of shedding light on the developmental and neurocognitive aspects of prosocial behaviour. It applies a battery of games to a population of children, teenagers and adult people I am currently analysing a set of data – already collected – of 100 subjects playing the Dictator/Investment game, the Ultimatum game and two test of cognitive control capabilities (the Go-Nogo and the MSIT), aiming to distinguish the neural activity and the different contribution of the three systems during development.

Finally, the fifth one is a book chapter that discusses evidence concerning the process of valuing others' perspectives (i.e., the weighing others' preferences to adapt our own behavior and achieve adequate social interaction) from an interdisciplinary viewpoint. First, we review economic research related to social decision-making, with emphasis on how the neoclassical economic theory has integrated other-regarding preferences into the decision-making algorithm. Then, we review how social skills develop, during lifetime, from identification to understanding others under the focus of social and developmental psychology and we finish discussing the neurobiological mechanisms underlying social skills and decision-making, highlighting how the cog-

nitive control and mentalisation neurocognitive systems interact to enable subjects to valuing others preferences. Our aim is to generate a starting point for building a more general explicatory bridge among the different disciplines that study complex human social behaviour.

2 Aspectos neurobiológicos y cognitivos del comportamiento social

Chapter to appear in the book *Comportamiento social de los invertebrados y vertebrados chilenos: mecanismos y función* (Luis Ebensperger and Antonieta Labra, Eds.), by:

Mauricio Aspé-Sánchez[*,1,2], Pablo Billeke[1] & Francisco Aboitiz[3,4]

[1] Laboratorio de Neurociencia Social y Neuromodulación (neuroCICS), Centro de Investigación en Complejidad Social (CICS), Facultad de Gobierno, Universidad del Desarrollo, Santiago, Chile.

[2] Instituto de Neurociencia, Universidad de Valparaíso, Valparaíso, Chile.

[3] Departamento de Psiquiatría, Escuela de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile.

[4] Centro Interdisciplinario de Neurociencias, Pontificia Universidad Católica de Chile, Santiago, Chile.

[*] Correspondencia a Mauricio Aspé-Sánchez. Email: maspes@udd.cl

2.1 Resumen

El presente capítulo busca esbozar – enfocándose en algunas especies de roedores y primates cuyos individuos forman vínculos afectivos– un marco explicativo para la existencia de algunos de los mecanismos evolutivos y neurocognitivos que permiten la formación de grupos sociales vinculados y, sobre estos grupos vinculados, el desarrollo de los sistemas neurocognitivos que están a la base de la expresión de dinámicas de cooperación dentro de sistemas sociales con atributos emergentes, como las que observamos en *Homo sapiens*.

Nuestro marco es pluralista, por lo que incluye mecanismos alternativos a presiones selectivas, en conjunto con condiciones ecológicas, del desarrollo, y procesos de evolución cultural. Además, considera que estos últimos procesos están a la base de la evolución de normas sociales, y que hacen plausible que algunas dinámicas de cooperación se sostengan en el tiempo mediante mecanismos de reciprocidad indirecta. Si bien nos detenemos con cierta profundidad en mecanismos neurogenéticos de la conducta social en algunos modelos roedores, luego ponermos el foco sobre *Homo sapiens*, para posteriormente describir brevemente algunas aves y roedores de la fauna local como potenciales modelos. Nuestra aproximación supone que el puzle neurocognitivo del comportamiento social es bidireccional: estructuras sociales complejas han influido sobre la evolución de un cerebro social y, a su vez, la evolución de un cerebro social ha posibilitado el desarrollo de estructuras sociales complejas. Si la relación entre estos fenómenos observados y las redes neuronales sobre los que se implementan es bidireccional y emergente, el corolario es que cerebros muy distintos, en especies muy distintas, pueden ser capaces de producir fenómenos de sociabilidad y socialidad similares.

2.2 Introducción

Tanto la sociabilidad (la tendencia de los individuos de una especie o población a formar grupos sociales discretos y relativamente permanentes; Kappeler y van Schaik 2002, Ebensperger y Hayes 2016) como la socialidad (la tendencia de un individuo a permanecer o buscar la presencia de otros de la misma especie; ver Réale et al. 2007) son fenómenos extendidos en el reino animal. Ambos han sido reportados en un amplio rango de especies, desde la conocida estructura social de las hormigas y las abejas (Gadagkar 1987, Wilson 1971) – pasando por otros invertebrados como la gamba (Synalpheus regalis; Duffy et al. 2000), hasta diversos vertebrados. En mamíferos, se han documentado más de 70 especies de roedores con conductas sociales (Beery 2019, Lacey y Sherman 2007); en cetáceos y primates, la gran mayoría de las especies muestran uno o más atributos sociales (Fox et al. 2017, Terborgh y Janson 1986), incluyendo *Homo sapiens* (Camerer 2003, Fehr y Fischbacher 2003). En Chile, en particular, existen al menos 296 especies que exhiben una o más formas de sociabilidad (Ebensperger, Capítulo 5).

En este capítulo, revisaremos algunos mecanismos evolutivos, neurobiológicos y cognitivos que subyacen a la formación de grupos sociales, con especial énfasis en las especies de roedores y primates cuyos individuos establecen interacciones afiliativas socio-positivas reiteradas en el tiempo: *i.e.*, individuos que forman lazos sociales (Sachser et al. 1998; Ostner y Schulke 2018) – a los que denominaremos también "vínculos", o "vínculos afectivos". Estos mecanismos permiten la expresión de rasgos de socialidad que van desde la crianza y la formación de lazos monógamos (Insel y Shapiro 1992, Young et al. 1999) hasta la manifestación de distintos niveles de cooperación a través de conductas tales como confianza (Camerer 2003, Fehr y Fischbacher 2002), altruismo (Jouventin et al. 2016, Fehr y Fischbacher 2003, Forsythe

et al. 1994), empatía (Frith y Frith 1999, Lamm et al. 2007), y reciprocidad directa e indirecta (Nowak y Sigmund 2005, Wilkinson 1988), dentro de estructuras sociales que presentan atributos emergentes.

Para esto, primero expondremos algunos requerimientos fisiológicos relacionados con la formación de grupos (Sección 7.2). A continuación, discutiremos el rol de los neuropéptidos sociales oxitocina (OXT) y arginina-vasopresina (AVP) en el establecimiento de lazos sociales (Sección 7.3). Luego revisaremos las principales teorías que se han desarrollado para explicar la evolución de un "cerebro social" (Sección 7.4 y 7.5) – entendido como el conjunto de áreas y redes cerebrales que están a la base de procesos involucrados en el procesamiento, almacenamiento y uso de información (generalmente sobre conespecíficos) en interacciones sociales (Billeke y Aboitiz 2013; Brothers 1990). A partir de esto, examinaremos algunas dinámicas de cooperación, utilizando como modelo un organismo que muestra elevados – y particulares – niveles de socialidad: *Homo sapiens* (Sección 7.6). Finalmente, abordamos el potencial de algunos animales chilenos (principalmente *Octodon degus*) como modelos para contrastar estas hipótesis (Sección 7.7).

Nuestra tesis central, esquematizada en la Figura 1, es que – satisfechas ciertas condiciones que hacen plausible la formación de agregados de individuos – los neuropéptidos sociales, al inhibir la actividad de la amígdala, disminuyen la ansiedad y la sensibilidad a la traición en contextos sociales, facilitando el establecimiento de lazos afectivos, el reconocimiento de conespecíficos y el acercamiento prosocial. Sobre la base de estos lazos, evidenciados principalmente en la relación cría-cuidador, redes neuronales relacionadas con procesos de alostasis son cooptadas, durante la evolución, por mecanismos de cognición social. En este contexto, procesos alométricos permiten el desarrollo de las cortezas de asociación – principalmente

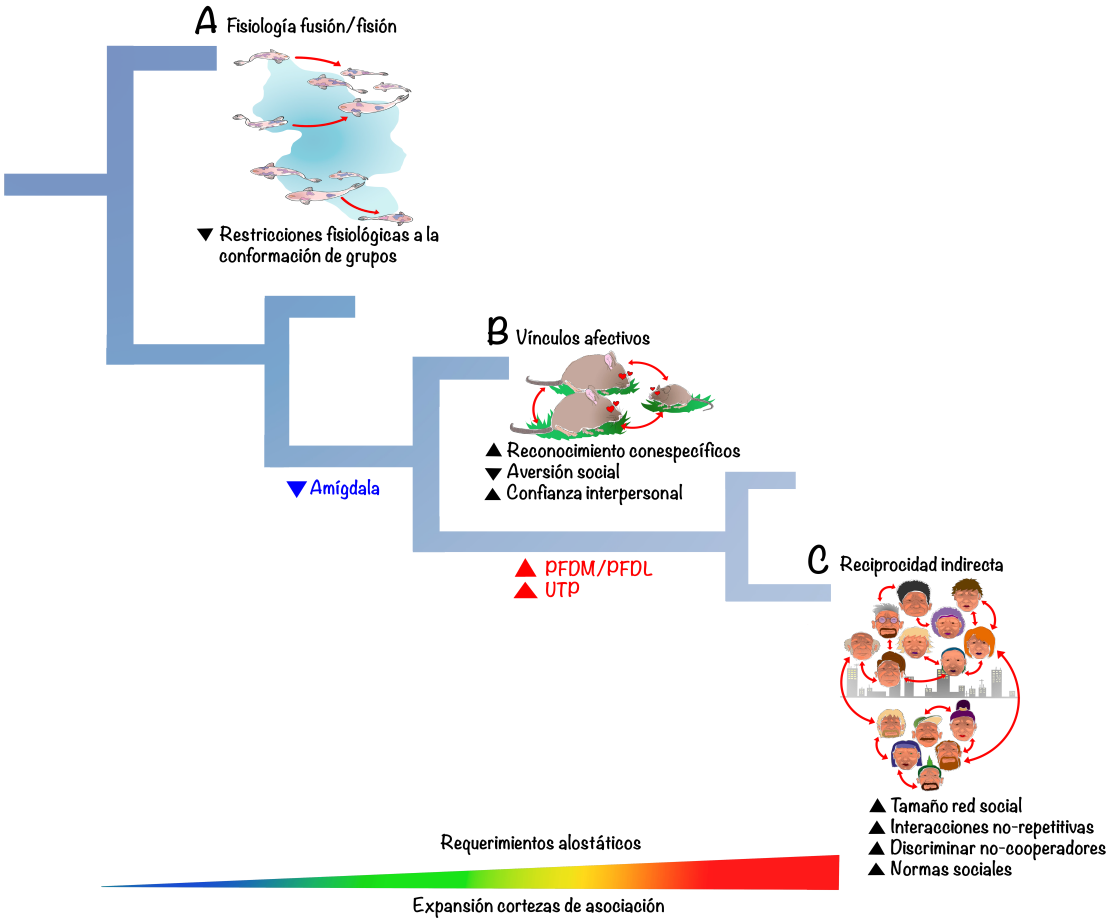


Figure 1: Evolución de las conductas sociales, desde la formación de grupos hasta el desarrollo de mecanismos de normas sociales y reciprocidad indirecta. **A.** Restricciones a la formación de agregados de individuos. Un primer requisito para la conducta social es que la relación entre las fisiologías individuales, sus ecologías y sus dinámicas evolutivas se mantengan dentro de parámetros que permitan la estabilidad del agregado. **B.** Los neuropéptidos sociales OXT y AVP facilitan el establecimiento de lazos afectivos y el acercamiento prosocial, putativamente disminuyendo la ansiedad social al modular la actividad de la amígdala. Los lazos sociales facilitan que redes neuronales relacionadas con procesos de alostasis implementen mecanismos de cognición social. **C.** La cortezas de asociación que emergen en este contexto – principalmente redes neuronales que involucran a la corteza prefrontal y la unión temporoparietal – implementan procesos de control cognitivo y mentalización, los que permiten la evolución de mecanismos de confianza, altruismo y reciprocidad directa e indirecta. Esto hace plausible la cooperación dentro de grupos grandes de individuos no relacionados genéticamente. En humanos, normas sociales (putativamente asociadas con la actividad de la corteza prefrontal dorsomedial) permiten el desarrollo de las complejas estructuras de cooperación que observamos en las sociedades actuales. [Ilustración: JC Aspé]

redes neuronales dominio-general que involucran a la corteza prefrontal y la unión temporoparietal – que subyacen a procesos de control cognitivo y mentalización, entre otros. Estos sistemas neurocognitivos, en conjunto con procesos de evolución cultural, permiten la evolución de normas sociales y mecanismos de reciprocidad indirecta, lo que hace plausible el desarrollo de las complejas estructuras de cooperación, entre grupos de numerosos individuos no relacionados genéticamente, que observamos en las sociedades humanas actuales.

Nuestra aproximación supone que la relación entre estos fenómenos observados y las redes neuronales sobre los que se implementan es bidireccional y emergente. Por esto, si bien el apartado se enfoca en *Homo sapiens*, los autores consideran que cerebros muy distintos, en especies muy distintas, pueden ser capaces de producir fenómenos de sociabilidad y socialidad similares.

2.3 Preámbulo: restricciones fisiológicas a la formación de grupos

Muchas de las complejas conductas sociales observadas en la naturaleza presentan atributos emergentes. Estos se definen, en el presente contexto, como las propiedades o características de un agregado de individuos que resultan (emergen) de las interacciones entre sus individuos constituyentes, y donde estas propiedades o características, a su vez, presentan propiedades causales sobre estos mismos individuos (Thompson y Varela 2001). En este sentido, la capacidad de formar agregados (o grupos) de conespecíficos es un primer requisito para el desarrollo del comportamiento social. Para que la socialidad exista tenemos, en primer lugar, que estar juntos.

La formación de grupos es un fenómeno bastante común en especies animales (Seebacher y Krause 2017, Krause et al. 2009), presentando costos y beneficios (una discusión detallada puede encontrarse en el Capítulo 5 de este mismo libro). Sus beneficios incluyen protección contra depredadores, mayor eficiencia de forrajeo y de intercambio de información, ventajas termorregulatorias, acceso a parejas reproductivas y ayuda para el cuidado de las crías (Seebacher y Krause 2017, Krause et al. 2009, Berry 2019). Dentro de sus principales costos se encuentran la competencia por alimento, parejas reproductivas y otros recursos; la transmisión de enfermedades, y también un aumento de la probabilidad de depredación (revisado en Lee 1994, Beery 2019, Ebensperger Capítulo 5).

El gregarismo depende, en sus niveles más básicos, de las características fisiológicas de los individuos que constituyen el grupo (Seebacher y Krause 2017). Estas características pueden favorecer, disminuir o incluso impedir la formación de grupos (Conradt y Roper 2000, Herbert-Read et al. 2011). Por ejemplo, una restricción a la formación de un cardumen es la hipoxia producida por la reducción del oxígeno disponible en el medio (Seebacher y Krause 2017). Además, el que un grupo se mantenga cohesionado requiere que los individuos sean capaces de sincronizar sus actividades, tal como ocurre durante las migraciones o el forrajeo – lo que se conoce como "teoría de sincronización de actividad" (Conradt y Roper 2000, Seebacher y Krause 2017). Por tanto, es crucial que la variabilidad de fisiologías no sea un impedimento para la cohesión grupal, provocando que fenotipos extremos rompan la cohesión, reduciendo así las ventajas que podrían obtenerse del gregarismo (Couzin y Laidre 2009, Seebacher y Krause 2017).

Cómo las fisiologías individuales podrían ser un impedimento para la cohesión grupal puede apreciarse en la interacción entre las variables *(i)* costos metabólicos y *(ii)* ca-

pacidad locomotora: dentro de un grupo, los individuos con máxima y mínima velocidad locomotora se comportan de modo subóptimo en cuanto a sus costos energéticos de locomoción y sus umbrales individuales de fatiga para mantener la cohesión grupal (Seebacher y Krause 2017). Si la magnitud de estas diferencias es muy elevada, estas pueden llegar a fisionar el grupo, o a forzar la formación de grupos distintos compuestos por individuos con fisiologías similares (Couzin y Laidre 2009, Seebacher y Krause 2017). Sin embargo, la variabilidad fisiológica puede aumentar la resiliencia de los individuos del grupo ante, por ejemplo, cambios en el ambiente (Ebensperger Capítulo 5, Seebacher y Krause 2017). En este sentido, existe una relación bidireccional entre las dinámicas de fisión y fusión, donde los organismos se agregan y desagregan en grupos de distinto tamaño, por un lado, y sus ecologías (Couzin y Laidre 2009, Seebacher y Krause 2017) y sus dinámicas evolutivas – tales como sus tasas de flujo genético y especiación –, por el otro (Kurvers et al. 2014, Farine et al. 2015). Por ejemplo, condiciones ecológicas como la disponibilidad de alimento influyen en cómo las diferencias fisiológicas determinan los eventos de fisión-fusión (Couzin y Laidre 2009, Seebacher y Krause 2017), mientras que la composición del grupo (cardúmenes más pequeños o más grandes) puede retroalimentar la fisiología de los individuos, alterando también los flujos genéticos y la selección (Seebacher y Krause 2017). De este modo, para que se formen agregados de individuos – un primer requisito para la existencia de la conducta social –, las interacciones entre las fisiologías de los individuos constituyentes, sus ecologías, y sus dinámicas evolutivas deben mantenerse dentro de parámetros que permitan la estabilidad del agregado.

La satisfacción de los requerimientos fisiológicos que posibilitan la formación de agregados de individuos, en conjunto con la evolución de neuropéptidos sociales, permiten la aparición no solo de grupos de conespecíficos sino que, además, de grupos de cone-

específicos que manifiestan vínculos. En el siguiente apartado, revisamos el rol de estos neuropéptidos, *i.e.* OXT y AVP, en fenómenos sociales que van desde el apego y la monogamia hasta la expresión de confianza, altruismo y empatía. Como veremos, OXT y AVP facilitan el establecimiento de vínculos afectivos, el reconocimiento de conespecíficos y el acercamiento prosocial, putativamente modulando la actividad de la amígdala.

2.4 Neuropéptidos sociales: establecimiento de lazos

La elevada variabilidad observada en los diferentes componentes de los sistemas sociales, como el vivir en grupos monógamos o polígamos, el tipo de cuidado parental, o la existencia de preferencias por individuos genéticamente relacionados, sugieren que estas diferencias están asociadas a mecanismos neurobiológicos igualmente diferentes (Couzin y Laidre 2009, Seebacher y Krause 2017). Sin embargo, existen mecanismos subyacentes a estos componentes que pueden mostrar un sorprendente grado de conservación, en un amplio espectro de grupos taxonómicos (Beery 2019, MacDonald y MacDonald 2010, Caldwell et al. 2008, Insel 2010, Donaldson y Young 2008). En esta línea, dos neuropéptidos – denominados "prosociales" – han sido involucrados en el desarrollo de estos lazos: OXT y AVP. Ambos neuropéptidos modulan la agresividad y la aversión social, lo que genera las condiciones para que establezcamos grupos de individuos más grandes, complejos, y con vínculos afectivos.

OXT y AVP son dos neuropéptidos muy conservados filogenéticamente, con un origen que se remonta a más de 700 millones de años (MacDonald y MacDonald 2010, Caldwell et al. 2008, Insel 2010). Poseen homólogos en diversos clados (Donaldson y Young 2008), estando presentes desde gusanos (*Caenorhabditis elegans*) a mamíferos (Beery 2019, Garrison et al. 2012, Althammer et al. 2018). En mamíferos, OXT

y AVP son sintetizadas en los somas de dos grupos neuronales localizados en los núcleos paraventricular y supraóptico del hipotálamo (Figura 2A). A continuación ambos neuropéptidos – que poseen una estructura química casi idéntica (Figura 2B) – son transportados a través de axones que proyectan a la pituitaria posterior, desde donde son liberados a la circulación (Loup et al. 1991, Meyer-Lindenberg et al. 2011, Zink et al. 2012). Además de actuar como hormonas en sus blancos periféricos (enumerados en la Figura 2C y D), también actúan dentro del cerebro como neuromoduladores y neurotransmisores, principalmente mediante liberación dendrítica (Landgraf y Neumann 2004, Leng y Ludwig 2008).

Los receptores de OXT y AVP (OXTR y AVPR, respectivamente) pertenecen a la superfamilia del receptor transmembrana G. En humanos, si bien se conoce solo una forma de OXTR, existen tres subtipos de AVPR, denominados AVPR1a, AVPR1b y AVPR2 (Ebstein et al. 2012, Loup et al. 1991, Meyer-Lindenberg et al. 2011). Dentro del cerebro, OXTR está expresado en la amígdala basolateral, el hipotálamo anterior y ventromedial, el núcleo olfatorio, la banda diagonal de Broca, el núcleo septal y la corteza cingulada anterior; en el caso de AVPR, esta se expresa en el núcleo septal, el tálamo y el núcleo amigadaloide basal (Meyer-Lindenberg et al. 2011 Aspé-Sánchez et al. 2016). Se considera que es a través de esta vía (liberación dendrítica de ambos péptidos actuando sobre blancos en el cerebro) que tanto OXT como AVP regulan conductas y cogniciones sociales tales como apego (Insel 2010, Young et al. 1999), exploración y reconocimiento social (Meyer-Lindenberg 2011), afiliación (Insel 2010), agresión (Berry et al. 2008), apareamiento (Young et al. 1999, Caldwell et al. 2008), comportamientos de tipo ansioso (Landgraf y Neumann 2004, Hammock et al. 2005) y, como veremos más adelante, probablemente altruismo (Knafo et al. 2005), empatía (Luo et al. 2015) y confianza interpersonal (Kosfeld et

al. 2005), entre otros (revisado en Aspé-Sánchez et al. 2016).

A principios de los años 90, Insel y Shapiro (1992) mostraron que la distribución de AVPR difiere en el cerebro de dos especies de topillos filogenéticamente cercanas, presentes en norteamérica: *Microtus ochrogaster* (topillo de las praderas), una especie caracterizada por un sistema de apareamiento monógamo, y *Microtus montanus* (topillo de la montaña), una especie con un sistema de apareamiento relativamente promiscuo (Young et al. 1999, Sadino y Donaldson 2018). Estos estudios revelan que, en cautiverio, inyecciones intracerebroventriculares de AVP aumentan la conducta de afiliación en el topillo monógamo de la pradera, pero no en el topillo promiscuo de la montaña (Young et al. 1999, Sadino y Donaldson 2018). Además, ratones transgénicos que expresan la variante de AVPR del topillo monógamo muestran un mayor comportamiento de afiliación en respuesta a inyecciones de AVP, lo que no ocurre al expresar en ratón la variante de AVPR del topillo promiscuo (Young et al. 1999). Respecto a OXT, Nakajima et al. (2014) demostraron que el bloqueo de la neurotransmisión en interneuronas de la corteza prefrontal medial (específicamente, en las interneuronas de somatostatina, las cuales, en ratones, expresan OXTR) provoca que las hembras pierdan interés en los machos durante la fase sexualmente receptiva del ciclo estral (Nakajima et al. 2014, Aspé-Sánchez et al. 2016). Los ratones nulos (también denominados *knock-out*) para OXTR – *i.e.*, ratones a los cuales se les ha inactivado el gen para OXTR – muestran déficits en comportamientos sociales, tales como mayor agresión y menor tiempo de exploración de conespecíficos, los que se revierten ante inyecciones intracerebroventriculares del péptido (Sala et al. 2015). En otros mamíferos, específicamente en ovejas, la liberación de OXT por estimulación vaginocervical durante el parto induce conductas de crianza y facilita el lazo madre-cría, putativamente alterando la actividad de neurotransmisores dentro

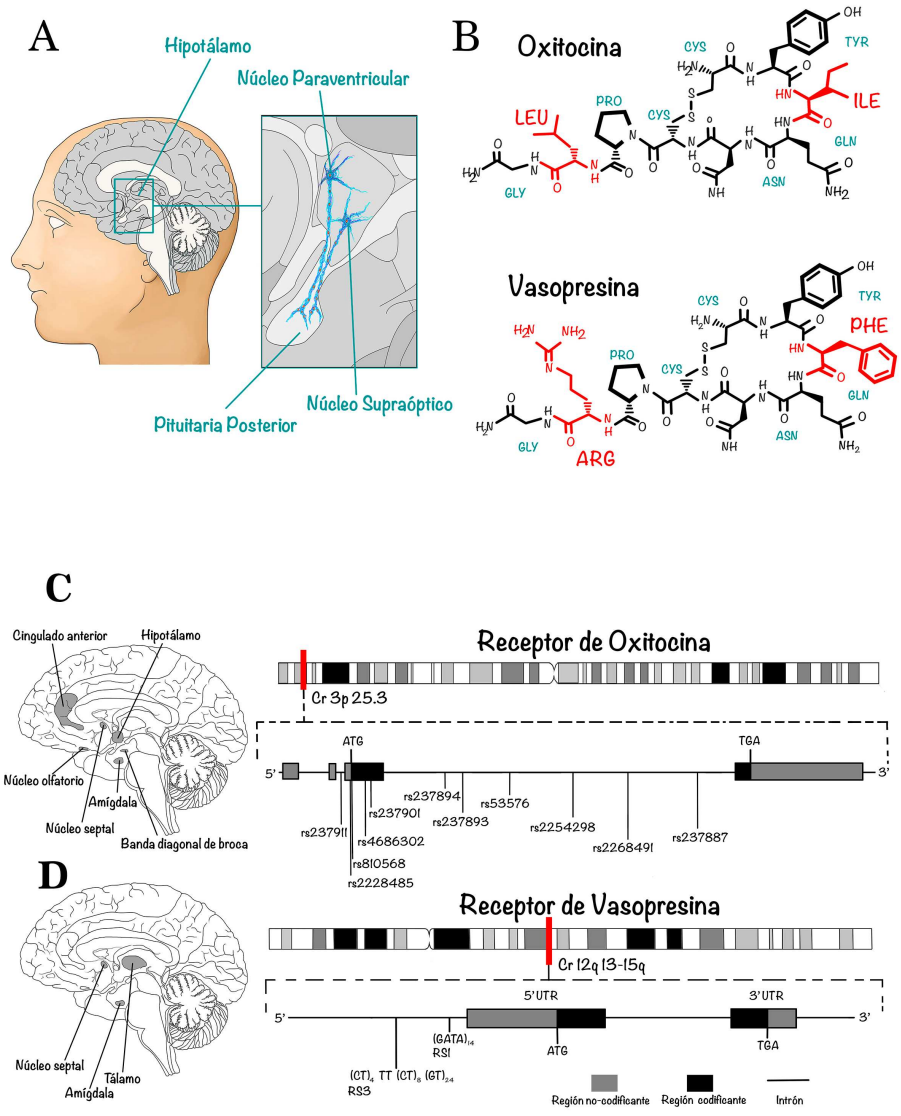


Figure 2: Los neuropéptidos sociales, OXT y AVP. **A.** Liberación periférica (pituitaria posterior) y liberación central (núcleos paraventricular y supraópticos). **B.** En mamíferos, ambos péptidos se diferencian solo en dos aminoácidos (en rojo): una isoleucina por una fenilalanina y una leucina por una arginina, en la tercera y octava posición, respectivamente. **C** y **D.** Esquemas de cromosomas mostrando la distribución de los receptores (izquierda) y estructura del gen (derecha) de OXT (cromosoma 3; arriba) y AVP (cromosoma 12, abajo). La barra roja representa la ampliación de un segmento de cromosoma, destacando la ubicación de sus polimorfismos. De especial interés son el polimorfismo de nucleótido único rs53576, en el caso de OXTR, y los microsatélites RS1 y RS3, en el caso de AVP. [Ilustración: JC Aspé. Modificado con permiso de Aspé-Sánchez et al., 2016]

del bulbo olfatorio, de modo tal que la madre responde selectivamente al olor de la cría (Broad et al. 2006).

Los hallazgos de Insel y Shapiro (1992) y Young et al. (1999), mencionados más arriba, fueron los primeros en reportar que la expresión diferencial del receptor de un neuropéptido puede resultar en diferencias en el grado de afiliación hembra-macho en dos especies de mamíferos. Además, muestran que tanto los polimorfismos genéticos – i.e., la ocurrencia de dos o más variantes en la estructura de un mismo gen, afectando la estructura y/o funcionalidad de la proteína o secuencia de ARN resultante – como el patrón espacial de expresión de estos receptores presentan variaciones especie-específicas, incluso en especies estrechamente relacionadas (Hammock y Young 2004, 2005). Esto sugiere que los resultados en roedores y otros mamíferos no son fáciles de extrapolar a otras especies. Por ejemplo, Fink et al. (2006) han postulado que, en primates, los polimorfismos de AVPR1a asociados a lazos sociales son evolutivamente distintos a los presentes en los roedores. A pesar de estas dificultades (y de que los estudios sobre el rol de estos polimorfismos en el comportamiento social humano han sido poco informativos; ej., Meyer-Lindenberg et al. 2011), existe evidencia que muestra que el gen para OXTR, y que al menos dos de los tres genes para AVPR, presentan polimorfismos característicos que podrían ser parte de las bases genéticas de la heterogeneidad de los rasgos sociales desplegados por *Homo sapiens*. Al parecer, existen mecanismos neurogenéticos subyacentes a la conducta social que son comunes a muchos mamíferos, *H. sapiens* incluido.

En humanos, la confianza, el altruismo y la reciprocidad han sido abundantemente estudiadas desde la teoría de juegos conductual (Camerer 2003). Confianza y reciprocidad han sido operacionalizadas utilizando, principalmente, un diseño experimental denominado "Juego de las inversiones" (Berg et al. 1995). En un Juego de las inver-

siones – y en su versión de decisiones binarias, el "Juego de las confianza" (McCabe et al., 2003) – participan dos jugadores, denominados Jugador 1 y Jugador 2. Ambos interactúan (usualmente) solo una vez, de forma anónima, y en conocimiento de todas las reglas del juego. El experimentador le entrega un monto de dinero $T al Jugador 1, quien puede manifestar confianza si decide enviarle alguna suma de su total disponible ($T) al Jugador 2. La suma enviada – digamos, $X – es multiplicada por un factor de intercambio (usualmente 3), y entregada al Jugador 2. El Jugador 2, por su parte, puede manifestar reciprocidad devolviéndole al Jugador 1 alguna suma del dinero – mayor a $0 y menor o igual a su total disponible, $3X – que el Jugador 1 le ha confiado (Berg et al 1995). El altruismo, por su parte, ha sido operacionalizado mediante el "Juego del dictador", donde el Jugador 1 recibe del experimentador una suma $T y, de modo anónimo, tiene la opción de manifestar altruismo al enviarle alguna suma de dinero – mayor a $0 y menor o igual a $T – al Jugador 2 (Forsythe et al., 1994). En ambos juegos, los jugadores siempre tienen la opción de no manifestar preferencias sociales – *i.e.*, comportarse maximizando exclusivamente su beneficio monetario personal, como es predicho por la teoría económica neoclásica (ver, por ejemplo, Stiglitz y Walsh 1993, Fehr y Fischbacher 2002, Fehr y Fischbacher 2003, Weintraub 2007) – si deciden enviar una suma igual a $0. Esta predicción es inconsistente con un abundante corpus de resultados experimentales (ver Johnson y Mislin 2011, Cox 2004, Henrich et al. 2004, Camerer 2003, Fehr y Fischbacher 2002).

En humanos, estas conductas han sido asociadas a polimorfismos en OXTR y AVPR. La mayoría de los polimorfismos en el gen OXTR corresponden a polimorfismos de nucleótido único, siendo rs53576 (cambio de G a A dentro del tercer intrón de OXTR) uno de los más estudiados (Figura 2C). En el caso de AVPR1, los polimorfismos más

estudiados son los microsatélites RS3 y RS1 (Meyer-Lindenberg et al. 2011; Figura 2D). En uno de los primeros estudios en humanos, Knafo et al. (2008) clasificaron los RS3 en AVPR1a de los participantes en versiones cortas (308 - 325 pares de base) y largas (327 - 343 pares de base), encontrando que los sujetos con la versión corta manifiestan menores niveles de altruismo – medido a través del Juego del dictador, en comparación con los sujetos con las versiones largas del alelo. Además, las repeticiones largas de RS3 se asociaron con niveles de ARNm en hipocampo más altos que las repeticiones cortas, *postmortem* (Knafo et al. 2008). En esta misma línea, Krueger et al. (2012) reportaron que individuos homocigotos para el alelo rs53576 G de OXTR exhiben un comportamiento de mayor confianza interpersonal – medido a través del Juego de la confianza, comparado con portadores del alelo A (Krueger et al. 2012). Estudios similares sugieren un rol de estos polimorfismos en otro aspecto de la socialidad humana: sujetos homocigotos para este mismo alelo exhiben mayor empatía, en comparación con los portadores del alelo A (Rodrigues et al. 2009).

Para entender los mecanismos mediante los cuales estos polimorfismos inciden en la conducta social se requiere conocer sus consecuencias funcionales. Sin embargo, la información disponible al respecto es escasa. Algunos reportes han asociado polimorfismos inter e intraespecíficos (Knafo et al. 2008, Young et al. 1999) con actividad promotora (Tansey et al. 2011), patrones de expresión cerebral de los receptores (Hammock et al. 2005, Knafo et al. 2008, Young et al. 1999), y regulación transcripcional (Hammock y Young 2005, 2004). Crucialmente, algunos polimorfismos parecieran estar relacionados con diferencias en el volumen de la amígdala y la corteza cingulada anterior: humanos homocigotos para el alelo G de rs2254298 en OXTR muestran un volumen de amígdala bilateral más pequeño, en comparación con los

portadores del alelo A (Inoue et al. 2010). En esta misma línea, los homocigotos para el alelo G en rs56579 presentan volúmenes más pequeños tanto de amígdala bilateral como de corteza cingulada anterior, también en comparación con los portadores del alelo A (Furman et al. 2011).

En resumen, los estudios anteriores destacan el rol de OXTR y AVPR en la formación de lazos sociales, y además sugieren que estos neuromoduladores permiten la expresión de diversos componentes de la socialidad. Si bien la información sobre los mecanismos mediante los cuales estos neuropéptidos influyen en la conducta social es aún insuficiente, existen estudios que correlacionan sus polimorfismos con volumen (y activación) de la amígdala – donde excitan diferentes subpoblaciones neuronales y modulan tareas sociales como discriminación de emociones (Huber et al. 2005, Ferretti et al. 2019) – y la corteza cingulada anterior, relacionadas con memoria aversiva y control cognitivo, respectivamente. Como veremos en la Sección 7.6, la acción de OXT sobre estas estructuras podría ser el mecanismo mediante el cual este neuropéptido disminuye la aversión social, facilita el reconocimiento de conespecíficos y aumenta la tendencia a confiar en ellos.

Sin embargo, esto no basta para explicar la capacidad de algunos organismos para formar sistemas sociales complejos – *i.e.*, estructuras sociales que presentan atributos emergentes. En la siguiente sección revisaremos las teorías desarrolladas para explicar la evolución de los sistemas neurocognitivos que posibilitan la formación de estas estructuras.

2.5 Más allá de los lazos: cerebros y sistemas sociales complejos

Como discutiremos en el presente apartado, en las últimas décadas se ha documentado una correlación entre el tamaño (relativo) del cerebro y la capacidad de distintas especies de primates para formar grupos sociales complejos (Dunbar 1998, Dunbar y Shultz 2007). Estos grupos sociales presentan atributos aditivos y/o emergentes, tales como *(i)* modularidad social jerárquica, donde unidades sociales de "bajo-orden" (como las familias) se anidan dentro de unidades sociales cada vez más grandes (fenómeno observado – además de en humanos – en babuinos, ballenas y elefantes) (Morrison et al. 2019), *(ii)* intrincadas redes de cooperación mediante reciprocidad (Nowak y Sigmund 2005), *(iii)* formación de coaliciones (Dunbar y Shultz 2007), *(iv)* estrategias como el engaño táctico (Byrne y Corp 2004) y, en el caso de especies que desarrollan sistemas culturales, *(v)* fenómenos como la interacción gen/cultura y la evolución cultural (ver Sección 7.5), *inter alia*. Sin embargo, a pesar de lo documentada que se encuentra la correlación entre tamaño (relativo) del cerebro y complejidad del grupo social, no existe consenso respecto a la importancia de los diversos factores planteados como causas tanto de la sociabilidad como de la evolución del cerebro primate.

Dentro de las interrogantes aún no resueltas podemos encontrar *(i)* cómo individuos con la capacidad para formar lazos sociales llegan a desarrollar sistemas sociales emergentes, y *(ii)* qué requerimientos neurocognitivos son necesarios para manifestar las conductas necesarias para la formación de estos sistemas (Sección 7.6). Otros fenómenos por dilucidar incluyen *(iii)* qué tipo de presiones evolutivas explican el aumento del tamaño cerebral y el establecimiento de patrones específicos de conec-

tividad neuronal que posibilitan *(i)* y *(ii)*. En este punto nos enfocaremos en la presente sección y en la siguiente.

Al menos en primates, buena parte de la discusión se ha enfocado en determinar si el papel preponderante en la evolución del cerebro lo han tenido condiciones ecológicas, tales como cambios en la dieta (González-Forero y Gardner 2018; DeCasien et al. 2017), factores sociales, tales como interacciones dentro de grupos cada vez más grandes y complejos (Dunbar 1998), o factores culturales, como el enseñar y aprender de conespecíficos (González-Forero y Gardner 2018). Además, se ha realizado un gran esfuerzo por dilucidar los mecanismos genéticos específicos asociados al desarrollo de cerebros sociales (Suzuki et al. 2018, Dunbar y Shultz 2017, Nithianantharajah et al. 2013, Tinbergen 1963). Actualmente, las explicaciones genéticas se enfocan principalmente en identificar los genes y mecanismos neurogenéticos relacionados con la acelerada evolución del cerebro dentro del linage humano (Nithianantharajah et al. 2013, Dunbar y Shultz 2017). Un ejemplo de ello es la duplicación del gen NOTCH2NL: screenings a eventos de duplicación génica en homínidos han reportado 35 duplicaciones específicas a *Homo sapiens*: dentro de ellas aparece, notablemente, NOTCH2NL, cuya vía de señalización aumenta la producción de células progenitoras corticales, lo que provoca un aumento del tamaño cortical (Suzuki et al. 2018). Otros genes relevantes incluyen PAX6 – gen altamente conservado, involucrado en el desarrollo de los ojos y el sistema nervioso central (Callaerts et al. 1997, Hanson et al. 1994), HEDGEHOG – responsable de la diferenciación de la placa y el tubo neural (Litingtung y Chiang 2000), y FOXP2 – asociado a desórdenes del habla y el lenguaje (Lai et al. 2001, Aboitiz 2017).

Como mencionamos, además de los mecanismos específicamente genéticos, se han planteado explicaciones ecológicas, sociales y culturales que buscan dilucidar qué

factores selectivos dan cuenta del tamaño y la organización del cerebro primate. Las explicaciones ecológicas se enfocan principalmente en las demandas requeridas por la búsqueda de alimento, donde se ha reportado que el tamaño del cerebro correlaciona con el desarrollo de nuevas técnicas de forrajeo, tanto en aves como en primates (Reader y Laland 2002, Dunbar y Shultz 2017). Existen estudios que muestran que, en primates, las especies frugívoras (cuya dieta es, cognitivamente, más demandante que una dieta folívora, por ser menos predecible espacial y temporalmente) presentan un mayor tamaño cerebral que las especies folívoras (DeCasien et al. 2017). Sin embargo, otros estudios reportan que no existe una relación entre la proporción de frutos en la dieta y el tamaño del cerebro cuando se incluye el tamaño del grupo social como variable control (Dunbar y Shultz 2017, Navarrete et al. 2016). Además, el desarrollo de estrategias específicas de forrajeo está ausente en la mayor parte de las especies animales, a pesar de la variabilidad en sus tamaños cerebrales. Por esto, las estrategias de forrajeo parecieran ser más una variable categórica que una continua (Reader y Laland 2002), y su asociación con el tamaño cerebral reducida únicamente a primates (Dunbar y Shultz 2017, Shultz et al. 2011).

Entre las explicaciones sociales, existen diversas hipótesis que intentan explicar el tamaño del cerebro primate. Estas son las de maquiavelismo, inteligencia vygostkiana, selección sexual y las basadas en la hipótesis del cerebro social. La hipótesis de maquiavelismo propone que la frecuencia de engaño táctico en distintas especies correlaciona con el volumen relativo de la neocorteza (Byrne y Whiten 1992, Byrne y Corp 2004). Sin embargo, esta hipótesis no explica la sociabilidad en primates, dado que la conducta maquiavélica, intrínsecamente competitiva, tendería a fragmentar los grupos sociales (Dunbar y Shultz 2017, Byrne y Whiten 1992), a menos que a este proceso se le opongan fuertes presiones selectivas. En este sen-

tido, la conducta maquiavélica sería más una consecuencia que una causa de vivir en grupos sociales (Dunbar y Shultz 2017). No obstante, aún si lo consideramos más como consecuencia que como causa, el engaño táctico podría constituir una presión selectiva para organismos que viven en grupos sociales donde el maquiavelismo, de hecho, existe (Whiten y Byrne 1997, Byrne y Whiten 1992).

La hipótesis social vygotskiana plantea que el cerebro humano se desarrolla porque las sociedades humanas son intrínsecamente cooperativas (Moll y Tomasello 2007). Sin embargo, dado que es una hipótesis restringida a las sociedades humanas, es difícil utilizarla para explicar diferencias cualitativas entre especies (Dunbar y Shultz 2017).

La hipótesis de selección sexual, por su parte, propone que el tamaño del cerebro aumenta como consecuencia de los mayores requerimientos neurocognitivos exigidos por un contexto donde las cópulas extra-pareja serían frecuentes (Miller 1999). Esta hipótesis es inconsistente con el hecho de que el tamaño del cerebro no está asociado con ninguno de los dos índices mejor establecidos de selección sexual: el tamaño relativo de los testículos y el grado de promiscuidad femenina (Schillaci y Grant 2006, Dunbar y Shultz 2017). Sin embargo, la selección de lazos de pareja (quizá un subconjunto de los fenómenos involucrados en los procesos de selección sexual) podría ofrecer pistas respecto a la evolución de cerebros grandes – al menos en mamíferos monógamos y aves (Beauchamp y Fernández-Juricic 2004, Pitnick et al. 2006), dado que las especies de aves que mantienen lazos monógamos de por vida tienen cerebros más grandes que las especies donde machos y hembras cambian de pareja anualmente (Dunbar y Shultz 2010, Shultz y Dunbar 2007).

La principal hipótesis social es la del cerebro social (Dunbar 1998, Dunbar y Schultz 2007). Esta propone que el aumento del tamaño de la neocorteza es producto de

las presiones selectivas asociadas al establecimiento de grupos sociales, dado que el cerebro debe responder a las altas demandas neurocognitivas requeridas por el vivir en grupos sociales complejos (Dunbar 1998, Dunbar y Schultz 2017). La asociación entre el tamaño del cerebro y la complejidad del grupo social parece ser un resultado consistente: por ejemplo, se han encontrado diferencias individuales entre el tamaño de la red social y el volumen de regiones frontales y temporales relacionadas con el cerebro social (Dunbar y Schultz 2007), como las revisadas en las secciones 7.5 y 7.6. Además, en mamíferos, el tamaño del cerebro correlaciona con indicadores de complejidad social, como la formación de coaliciones (Dunbar y Schultz 2007), y con estrategias sociales tales como engaño táctico, en primates (Byrne y Corp 2004).

Finalmente, la hipótesis de inteligencia cultural propone que cerebros más grandes permiten la transmisión de información a través de imitación o mímica (Reader y Laland 2002, Reader et al. 2011, van Schaik et al. 2012). Sin embargo, en el registro arqueológico, las competencias técnicas y la transmisión cultural parecen asemejarse más a cambios cualitativos – incluyendo largos periodos de estasis (Gould y Eldredge 1972) – que cuantitativos (Reader y Laland 2002). Esto sugiere que la hipótesis de inteligencia cultural podría ser útil para explicar la evolución del cerebro solo en una pequeña proporción de especies: los grandes simios y los humanos (Dunbar y Shultz 2017, Shultz et al. 2012). Además, existen reportes que sugieren que la estructura propia de los grupos sociales primates (redes multicapa con baja panmixia; King y Stanfield 1997), en lugar de acelerar, enlentece la tasa en la que las innovaciones se difunden dentro de un grupo (Dávid-Barrett y Dunbar 2012, Dunbar 2011). Por otra parte, el que los individuos aprendan de otros sus estrategias de forrajeo, su conocimiento técnico o sus reglas culturales no dice nada, por sí mismo, de por qué existen grupos con lazos sociales: quizá el uso de la información social disponible

sea más una exaptación (Gould y Vrba 1982) del vivir en grupos que una causa de ello (Dunbar y Shultz 2017). En este sentido, un fenómeno crucial a considerar es el desacople entre la evolución del cerebro y la dinámicas de desarrollo de entidades culturales, si tomamos en cuenta que la evolución cultural – que ha sido documentada en varias especies, incluyendo la marmota (*Mungos mungo*; Sheppard et al. 2018), no sigue las mismas dinámicas de la evolución biológica (ver Sección 7.5).

Las hipótesis revisadas en este apartado no están exentas de problemas. En primer lugar, algunos estudios se han enfocado en examinar la evolución de la arquitectura y el tamaño cerebral de los primates sobre la base de *una* explicación; generalmente, una de las enumeradas en los párrafos precedentes. Además, respecto a la hipótesis del cerebro social, intentos por establecer relaciones causales (y no simplemente correlacionales) a través de modelos que estiman el costo metabólico del cerebro, han mostrado que el tamaño relativo de éste se explica por una combinación de un 60% de factores ecológicos (*i.e.*, los factores no-sociales impuestos por la naturaleza, como cambios en la dieta, o el desarrollo de técnicas de forrajeo y procesamiento de comida; Clutton-Brock y Harvey 1980, Rosati 2017), y sólo un 30% de demandas dadas por interacciones cooperativas, sumadas a un 10% de demandas dadas por interacciones de competencia inter-grupal (González-Forero y Gardner 2018). Además, González-Forero y Gardner (2018) muestran que la competencia intragrupal posee un peso insignificante para explicar la evolución del cerebro humano, otorgándole más peso a factores culturales que sociales. Esto sugiere que explicaciones adaptacionistas (Orzack y Forber 2017, Gould y Lewontin 1979) – como algunas provenientes desde la psicología evolucionaria (Murphy 2003, Fodor 2001) – son insuficientes para explicar la evolución de la socialidad y el cerebro humano. Por otra parte, si bien estas correlaciones están bien documentadas, no entregan información respecto a si

los grupos sociales más grandes causan cerebros más grandes o si, por el contrario, son los cerebros más grandes los que permiten la existencia de grupos sociales más grandes (Powell et al. 2017, González-Forero y Gardner 2018). Finalmente, al menos en antropoides, existen reportes que muestran que es la reorganización cerebral, y no el tamaño relativo, lo que caracteriza su evolución (Smaers y Soligo 2013, Aboitiz 2017).

En resumen, los animales sociales parecieran tener cerebros más grandes pero, proporcionalmente, con una menor cantidad de neuronas, cuyas conexiones establecen redes neuronales con topografías distintas a las encontradas en cerebros de menor tamaño. Además, no está claro si el aumento en el tamaño (relativo) de la corteza es un antecedente o un consecuente de la conformación de grupos sociales. En la sección siguiente se discuten algunas de las propiedades de estas redes de neuronas, junto con otros mecanismos evolutivos involucrados, considerando que el fenómeno debe ser explicado apelando a una diversidad de mecanismos y procesos, entre ellos la evolución cultural.

2.6 Pluralismo en la evolución del cerebro social

La sección anterior describe diversas hipótesis acerca de cómo la evolución del cerebro y la conducta social primate podría ser el resultado de una variedad de presiones selectivas (Dunbar y Shultz 2007, Healy y Rowe 2007). Sin embargo, quedan interrogantes respecto a la importancia relativa de otras fuerzas evolutivas (en el sentido de Sober 2014), además de procesos de evolución cultural. Diversos estudios han resaltado el rol de estas fuerzas en fenómenos como la conformación del tamaño craneal y cerebral, y en el establecimiento de patrones de conectividad neuronal (Atzil et al. 2018, Schroeder y Ackermann 2017, Rubinov 2015, Buckner y Krienen 2013, Weaver

et al. 2007, Ackermann y Cheverud 2004).

Si bien existe polémica respecto a la definición de sociabilidad (Krause y Ruxton 2002, Ebensperger Capítulo 5), en términos generales podemos decir que es social una especie cuyos individuos, la mayor parte del tiempo, realizan sus actividades en presencia de conespecíficos (Ebensperger Capítulo 5). En una definición un poco más restringida – y siguiendo a Atzil et al. (2018), un animal social podría definirse como aquél que es inicialmente incapaz de sobrevivir por cuenta propia, por lo que requiere de conespecíficos para su supervivencia. Dicho de otro modo, requiere de otros para su alostasis, definida como el conjunto de procesos mediante los cuales los organismos se anticipan a sus requerimientos homeostáticos (Sterling 2012) – por ejemplo, cuando una cría busca la proximidad de su madre al comenzar a sentir hambre. Esto hace que los estímulos sociales sean salientes para estos organismos, los cuales aprenden a regular los procesos alostáticos propios y de sus conespecíficos mediante mecanismos de cognición social (Atzil et al. 2018, Atzil y Barrett 2017).

Existe evidencia que muestra que gran parte de los procesos neurocognitivos están causados por activaciones neuronales espacialmente distribuidas: el cerebro es un órgano que genera estados funcionales emergentes (Yuste 2011) a través de poblaciones (o "asambleas") neuronales que presentan coactivación espacio-temporal (Churchland y Sejnowski 1992): se ensamblan y desensamblan dinámicamente, dentro del orden de los milisegundos (Yuste y Tank 1996, Mesulam 1990, Felleman y van Essen 1991). Estas poblaciones distribuidas espacialmente, y con coactivación temporal, se denominan "redes neuronales". Las redes neuronales suelen ser bastante "'promiscuas", o "dominio-general": las mismas estructuras que se encargan de procesar una tarea dentro del dominio A muchas veces procesan también tareas dentro de los dominios B o C. Esta promiscuidad genera polémica dentro de la comunidad

de neurobiólogos cuando se intenta definir el rol específico de determinadas estructuras y redes neuronales, *i.e.*, cuando se intentan definir los "módulos" corticales (Fodor 1986) que subyacen al procesamiento de un dominio específico. Muchas veces esta modularidad, o dominio-especificidad, derechamente no existe (Spunt y Adolphs 2017). El módulo cerebral se encuentra en una crisis de paradigma (Fuster 2000), incluso por parte de sus principales exponentes (Fodor 2001).

Esto puede apreciarse en la Figura 3. En el presente contexto, las redes neuronales que subyacen a la conducta social se solapan con las que subyacen a la regulación alostática (Atzil et al. 2018, Barrett y Satpute 2013, Kleckner et al. 2017). Por ejemplo, la red neuronal de regulación alostática presenta estructuras cerebrales comunes con la red de modo por defecto (Kleckner et al. 2017), cuya actividad es mayor en reposo, cuando los sujetos no están involucrados en una tarea que requiere esfuerzo cognitivo (Fox et al. 2006, Power et al. 2011). La red de modo por defecto se solapa con la red de mentalización (Barrett y Satpute 2013, Raichle 2015, Raichle et al. 2001, Ramírez-Barrantes et al. 2019; ver Sección 7.6). Todas estas redes, a su vez, subyacen a procesos interoceptivos (Kleckner et al. 2017, Atzil et al. 2018). Esto sugiere que las redes neuronales que son cruciales para el desarrollo de la conducta social están implementadas sobre redes de mentalización, y que estas, a su vez, han cooptado estructuras relacionadas con interocepción y alostasis (Figuras 3 y 4).

La evolución de las redes neuronales recién mencionadas presenta características llamativas. Por ejemplo, la expansión de las cortezas de asociación – donde en primates se encuentran las redes de interocepción, alostasis, mentalización y modo por defecto (Atzil et al., 2018) – se habría producido por procesos alométricos (Shingleton 2010), generando una creciente distancia entre las cortezas sensoriales y motoras, a través de la filogenia (Buckner y Krienen 2013, Margulies et al. 2016). Como consecuencia

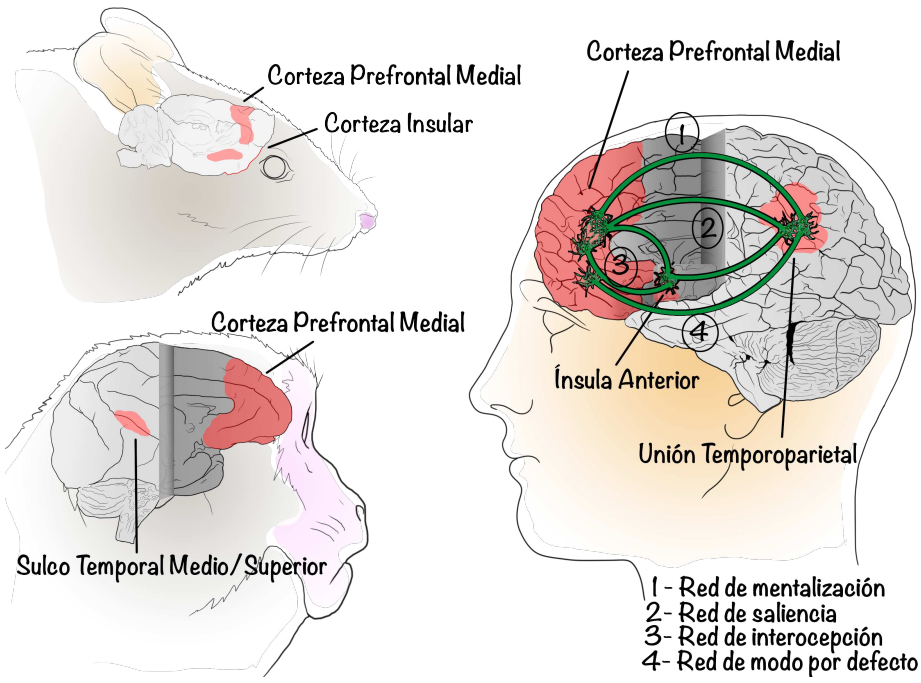


Figure 3: Estructuras y redes neuronales involucradas en procesamientos de saliencia, interocepción y modo por defecto en *Mus musculus*, *Macacus rhesus* y *Homo sapiens*. En este último, se incluye la red de mentalización. La red de mentalización, en humanos, incluye la corteza prefrontal medial y la unión temporoparietal (1). La red de saliencia (2) incluye la corteza prefrontal medial, la ínsula anterior y la unión temporoparietal. La red de interocepción (3) incluye la corteza prefrontal medial y la ínsula anterior. La red de modo por defecto (4) incluye la corteza prefrontal medial y la unión temporoparietal. PFM: corteza prefrontal medial; CI: corteza cingulada anterior; STM/STS: sulco temporal medio/superior; IA: ínsula anterior; UTP: unión temporoparietal. [Ilustración: JC Aspé]

de esto, los destinos celulares de las neuronas de la corteza de asociación están menos determinados por gradientes moleculares, dado que se van alejando paulatinamente de los puntos de anclaje molecular que emiten las señales que diferencian a los progenitores neuronales como neuronas somatosensoriales o motoras (Buckner y Krienen 2013). Esto es consistente con el hecho de que las redes neurales que subyacen al sistema de afiliación social se desarrollan durante la infancia, estando virtualmente ausentes en neonatos (Gao et al. 2016). Por ejemplo, la red de modo por defecto es

una de las primeras redes dominio-general en desarrollar su topología adulta característica: aún así, en neonatos humanos aparece como una única región aislada en la corteza cingulada posterior, adquiriendo su topología característica recién a los seis meses de edad (Gao et al. 2013, Gao et al. 2015, Atzil et al. 2018). En esta misma línea, existen reportes que muestran que, en humanos, la estructura de giros y sulcos dentro del cerebro presenta una menor determinación genética respecto a la estructura del cerebro en chimpancés (Gómez-Robles et al. 2015). Tanto el alto grado de plasticidad de estas redes neuronales durante la infancia, como también su posible desarrollo a partir de redes que subyacen a procesos alostáticos, muestran que difícilmente las redes involucradas en la socialidad humana vienen cableadas de nacimiento (Atzil et al. 2018), lo que las haría más plástica a los procesos de socialización cultural y a la experiencia, sobre todo durante el desarrollo temprano.

Es destacable que los cerebros más grandes tienen, proporcionalmente, una menor cantidad de neuronas (Aboitiz 2017, Herculano-Housel 2015, Haug 1987). Al menos en antropoides, existen reportes que muestran que es la reorganización cerebral, y no el tamaño relativo, lo que caracteriza su evolución (Aboitiz 2017, Smaers y Soligo 2013). En esta línea, estudios que han aplicado análisis de redes a datos de conectividad estructural y funcional muestran que los patrones de conectividad de las cortezas de asociación mencionadas anteriormente presentan propiedades de redes no-canónicas. En las redes canónicas, la conectividad entre estructuras es serial y jerárquica, con proyecciones de retroalimentación y anteroalimentación principalmente entre estructuras que se encuentran espacialmente cercanas y que conectan áreas disímiles – progresivamente, desde áreas sensoriales hacia cortezas de asociación, y luego a la corteza motora. En las redes con propiedades no-canónicas, en cambio, existe una propensión a proyectar conexiones hacia áreas similares, que

no presentan una estructura clara de retro ni anteroalimentación, y que están distanciadas espacialmente (Buckner y Krienen 2013, Selemon y Goldman-Rakic 1988, Goldman-Rakic 1988). Dentro de estas cortezas de asociación, la red de modo por defecto, por ejemplo, está situada a la mayor distancia geodésica de las cortezas motoras y sensoriales: de esta forma, esta red se encuentra "desacoplada" de estímulos sensoriales y motores inmediatos, lo que podría estar en la base de procesos tanto autorreferenciales como de mentalización (Margulies et al. 2016). Otros trabajos sugieren que los patrones de conectividad de estas redes podrían corresponder más a *spandrels* (ver Gould y Lewontin 1979) que a presiones selectivas propiamente tales. Por ejemplo, un estudio en los conectomas de *Drosophila melanogaster* y *Mus musculus* reporta que la topología de sus redes neuronales no se explica por restricciones debidas a presiones evolutivas sobre los costos de conectividad, sino que son subproductos de presiones estructurales – *i.e.*, *spandrels* (Rubinov 2015, Gould y Lewontin 1979). Otro estudio sugiere que la deriva génica (cambios en la proporción de genes o genotipos producidos únicamente por azar – "errores de muestreo" genéticos de una generación a la siguiente, sin participación de presiones selectivas; Millstein 2017) podría ser uno de los procesos evolutivos centrales que subyace a la evolución del neurocráneo en homínidos (Ackermann y Cheverud 2004). Esto es consistente con el hecho de que el tamaño de los grupos sociales en que vivían nuestros antepasados cazadores-recolectores era pequeño (50 - 100 individuos; Nowak y Sigmund 2005), los cual hace plausible que la deriva génica constituya una importante fuerza evolutiva.

Como mencionamos anteriormente, la reducida determinación genética en la estructura del cerebro humano, su alto grado de plasticidad durante la infancia, y la importancia de los procesos alostáticos en la conformación de sus redes neuronales muestran que el cerebro humano es muy sensible a la experiencia y a los procesos

de socialización cultural, sobre todo durante el desarrollo temprano. En esta línea, un punto crucial a destacar es el papel que puede tener la evolución cultural sobre el desarrollo de conductas sociales. Esto corre no solo para humanos, sino para diversas especies. Un estudio reciente en marmotas (*Mungos mungo*), por ejemplo, muestra cómo los procesos de transmisión cultural explican mejor la conducta de forrajeo, en comparación a los procesos de herencia genética (Sheppard et al. 2018). La evolución cultural también ha sido reportada en delfines y ballenas, utilizando la riqueza del repertorio social como variable proxy de complejidad cultural (Fox et al. 2017). En primates, capuchinos y gorilas – además de bonobos y chimpancés – evidencian procesos de transmisión cultural, juzgando por los tres mil años de uso de herramientas (con significativas variaciones en sus usos hace 2.400 y 300 años), en el caso de los capuchinos, y las reuniones anuales y estructuras de amistades de por vida, en el caso de los gorilas (Falótico et al. 2019; Morrison et al. 2019).

Los procesos culturales tienen impacto sobre la evolución genética, como lo muestra el caso paradigmático de coevolución gen-cultura (Feldman y Laland 1996) observado en la tolerancia a la lactosa (Beja-Pereira et al. 2003). Además, pueden explicar fenómenos sociales complejos, como algunos aspectos de la evolución del lenguaje, requiriendo solo la existencia de sistemas cerebrales de dominio general tales como (i) la adscripción de intenciones comunicativas a un interlocutor – *i.e.*, alguna capacidad de mentalización – y (ii) la capacidad de generar mapeos uno-a-uno entre significante y significado (Smith 2002). Finalmente, la teoría de construcción de nichos postula que la selección no actuaría a partir de un medio fijo que impone presiones selectivas sobre los organismos, sino que los organismos, activamente, construimos los nichos ecológicos – y socioculturales (Laland y O'Brien 2011, Boyd et al. 2011) – ante los cuales nos adaptamos (Laland et al. 2000). Esto es particularmente relevante

para organismos que desarrollan cultura (Laland et al. 2001, Boyd et al. 2011): diversos procesos culturales podrían crear y/o modificar las presiones selectivas a las que nos enfrenta el ambiente. Como corolario, a los procesos de selección natural que habrían moldeado sistemas dominio-específicos subyacentes al desarrollo de las capacidades sociales y lingüísticas (Aboitiz 2018) hay que añadir un sinnúmero de procesos de evolución cultural sobre los que pueden emerger dinámicas sociales complejas, sin requerir, necesariamente, una heredabilidad genética ni una arquitectura neural dedicada.

De este modo, las explicaciones para la evolución de un cerebro social deberían integrar explicaciones genéticas – que incluyen presiones selectivas, deriva génica, relajo en el control genético y subproductos estructurales, como *spandrels*; condiciones ecológicas – grupos pequeños que permitieron la deriva, pero con relaciones estrechas que generan presiones para establecer, por ejemplo, mecanismos de reputación (Sección 7.6; Nowak y Sigmund 2005); y del desarrollo – menor control sobre el destino de los progenitores celulares en las cortezas de asociación, y un aumento de la neotenia en el desarrollo del cerebro; además de procesos de construcción de nichos y de evolución cultural. Como corolario, el puzle neurocognitivo del comportamiento social es bidireccional, requiriendo responder cómo estructuras sociales complejas influyeron sobre la evolución de un cerebro social y, a su vez, cómo la evolución de un cerebro social posibilitó el desarrollo de estructuras sociales complejas.

En el siguiente apartado, expondremos los principales mecanismos neurocognitivos que están a la base de algunas de las dinámicas de cooperación observadas en diversas especies – incluidos algunos grandes simios, enfocándonos en *Homo sapiens*. Consideramos que, en animales gregarios donde los neuropéptidos prosociales han facilitado el acercamiento prosocial, estos mecanismos se implementan sobre redes neuronales

dominio-general de mentalización y control cognitivo – que incluyen estructuras como la unión temporoparietal (mentalización), lóbulo prefrontal ventromedial (control cognitivo) y lóbulo prefrontal dorsolateral (normas sociales) – y en cuya evolución es esencial considerar la importancia de factores no adaptativos, como el nicho cultural que permite el desarrollo de mecanismos de reputación y reciprocidad indirecta sobre la base de normas sociales.

2.7 Cerebros sociales: mecanismos neurocognitivos

En algunos organismos, los mecanismos de afiliación y sus sistemas neuronales subyacentes adquieren características que, más allá de simplemente permitir la formación de grupos con lazos sociales, sustentan mecanismos que permiten la formación de complejas dinámicas de cooperación dentro de grupos con atributos emergentes.

Estas interacciones existen dentro de redes sociales modulares y jerarquizadas, con unidades sociales de "bajo-orden" que se agrupan y anidan dentro de unidades cada vez más grandes (Morrison et al. 2019). En este contexto, un gran número de interacciones se tornan anónimas y no-repetitivas, y ocurren entre individuos que no están relacionados genéticamente (Nowak y Sigmund 2005) – lo cual es especialmente prominente en humanos (Wilson 1975, Trivers 1985, Hamilton 1996). Esto aumenta los incentivos para que quien inicia la cooperación sea traicionado por su contraparte, si ésta decide no cooperar (ver, por ejemplo, Berg et al. 1995, Coleman 2000, Camerer 2003, Nowak y Sigmund 2005). Por tanto, la cooperación depende fundamentalmente de preferencias sociales – *i.e.*, depende de la socialidad de sus participantes, tales como confianza, altruismo, y reciprocidad directa y/o indirecta (Nowak y Sigmund 2005, Fehr y Schmidt 2001, Camerer 2003).

En neurociencia social, se han identificado al menos dos redes neuronales dominio-general cuya actividad pareciera ser necesaria para la manifestación de estas preferencias (Figura 3 y 4): la red de control cognitivo (Holroyd y Coles 2002, Ullsperger et al. 2014) y la red de mentalización (Decety y Lamm, 2007). La red de control cognitivo permite diversos procesos neurocognitivos que guían la conducta (y los pensamientos) de acuerdo a las metas y planes del organismo (Posner y Snyder 1975), a cuya base se encuentra la actividad de la corteza prefrontal (Holroyd y Coles 2002, Shenhav et al. 2013, Ullsperger et al. 2014, Yamagishi et al. 2016). En este apartado nos centraremos en la red de mentalización, por el hecho de que en humanos – y quizá en grandes simios (Kano et al. 2019, Premack y Woodruff 1978) – pareciera sustentar los procesos neurocognitivos que están a la base de la capacidad de adscribir deseos, intenciones, creencias y agencia a otro organismo (Decety y Lamm 2007, Costa et al. 2008, Frith y Frith 1999).

La red de mentalización, representada esquemáticamente en la Figura 3, permite satisfacer un requisito fundamental para el desarrollo de la sociabilidad en humanos: la capacidad de discriminar entre personas y cosas (McCabe et al. 2001, Frith y Frith 1999). A través de la experiencia, esta red permite predecir y explicar la conducta de otro individuo, al cual le atribuimos cierto nivel de agencia (Billeke et al. 2017, Vogeley 2017, Ord y Martins 2010). Además, permite formar impresiones sociales con el fin de ajustar la conducta comunicativa de acuerdo al contexto social (MacLean et al. 2014). Como hemos mencionado anteriormente, el solapamiento de estructuras de la red de mentalización con las de la red de modo por defecto, la cual se solapa, a su vez, con la red de alostasis (Figura 3), podría arrojar luces sobre la evolución de las redes neuronales que están a la base de procesos de cognición social (ver sección 7.5).

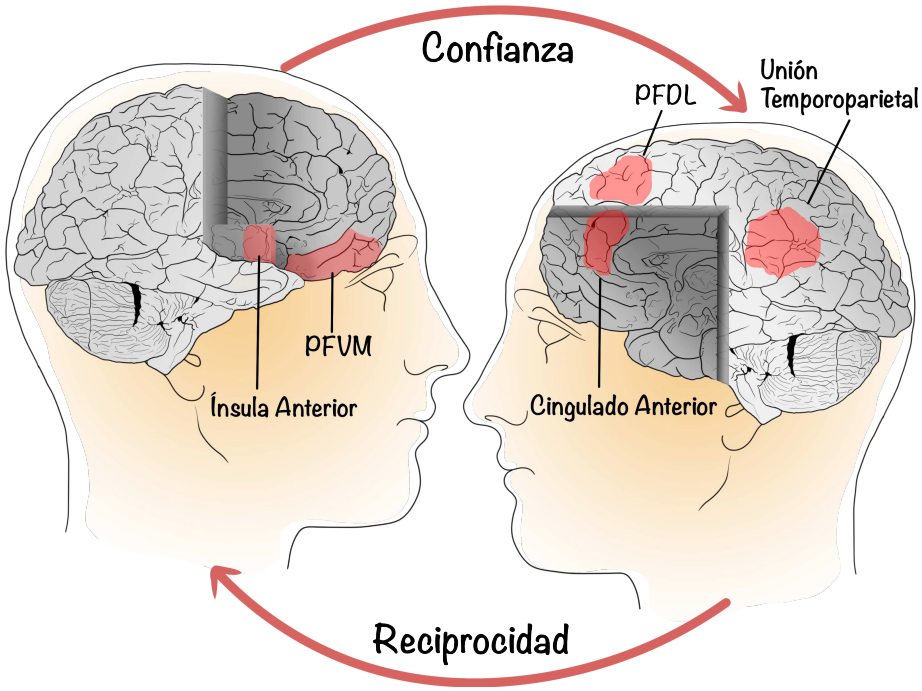


Figure 4: Redes neuronales implicadas en los procesos de confianza y reciprocidad directa e indirecta. PFVM: corteza prefrontal ventromedial; PFDL: corteza prefrontal dorsolateral. La ilustración no incluye el área septal, involucrada en el mantenimiento de dinámicas de confianza, posiblemente a través de la liberación de OXT y AVP. [Ilustración: JC Aspé]

En un experimento de mentalización típico, sujetos humanos se enfrentan a una tarea de falsa creencia (Samson et al. 2004). En esta tarea, los sujetos experimentales observan un vídeo o una historieta donde el personaje A esconde un objeto en un lugar X, y luego sale de escena. El objeto es posteriormente movido a un escondite Y por el personaje B, hecho que ocurre fuera del alcance visual del personaje A, pero que es presenciado por el sujeto experimental. Cuando el personaje A vuelve a entrar a escena, se le pregunta al sujeto experimental dónde A buscará el objeto. Si el sujeto experimental ha desarrollado la capacidad de mentalización (típicamente, sujetos humanos mayores de 4 años de edad; Soto-Icaza et al., 2019), responderá que el personaje A buscará el objeto en el lugar X, debido a que el sujeto experimental

es capaz de inferir que A mantiene una *falsa creencia* de que el objeto permanece allí. Un sujeto que no ha desarrollado la capacidad de mentalización responderá que A buscará el objeto en el lugar Y. Al medir la actividad neuronal de los sujetos mediante resonancia magnética funcional, diversos estudios han reportado una mayor activación de la (a partir de esto denominada) red de mentalización – notablemente la corteza prefrontal y la unión temporoparietal izquierda – cuando los sujetos resuelven con éxito esta prueba (Samson et al. 2004, Amodio y Frith 2006, Decety y Lamm 2007, Adolphs 2009). Sujetos con daño en la unión temporoparietal presentan déficits tanto al resolver esta tarea como en tareas donde deben procesar estímulos sociales relevantes, tales como la dirección de la mirada de otro organismo, o inferir conductas orientadas a objetivos (Samson et al. 2004).

El rol de las redes de mentalización y control cognitivo en la producción de una socialidad compleja puede ejemplificarse tomando como base experimentos que buscan elicitar confianza, altruismo y reciprocidad (Figura 4). Se ha determinado, mediante estudios con resonancia magnética funcional en sujetos que participan en el Juego de la confianza, que la opción de confiar en un conespecífico, al menos en humanos, está asociada a una mayor actividad de la red de mentalización: la corteza prefrontal ventromedial muestra mayor activación solo cuando los sujetos deciden confiarle dinero a otro humano, y no cuando le "confían" dinero a una lotería programada en un computador (McCabe et al. 2001, Rilling et al. 2002, Delgado et al. 2005, King-Casas et al. 2005, Krueger et al. 2007, Baumgartner et al. 2008). La unión temporoparietal, parte de esta misma red, muestra mayor activación cuando los sujetos reciprocan una acción de confianza que involucra un alto riesgo para quien confió – es decir, cuando los sujetos experimentales son recíprocos con conespecíficos que, al confiar, se arriesgan a perder una gran cantidad de dinero (van den Bos

et al. 2009). Otros estudios muestran que la unión temporoparietal, junto con la corteza prefrontal ventromedial, conforman una red que participa en el control de impulsos egoístas, aumentando su actividad cuando los sujetos manifiestan conductas altruistas (Hutcherson et al. 2015). La red de control cognitivo, necesaria para la inhibición del egoísmo y la toma de decisiones estratégicas y normativas, también está involucrada: la corteza cingulada anterior (Delgado et al. 2005, Shenhav et al. 2013, van den Bos et al. 2009) y la corteza prefrontal dorsolateral (Baumgartner et al. 2008, Yamagishi et al. 2016), muestran una mayor actividad cuando sujetos devuelven la confianza en una medida menor o mayor de lo que creen que espera quién confió en ellos, respectivamente (Chang et al. 2011; Figura 4). La actividad de la corteza prefrontal dorsolateral, como veremos más adelante, parece ser necesaria para el procesamiento de normas sociales (Yamagishi et al 2016).

Como se discutió anteriormente, el neuropéptido OXT se ha asociado con conductas prosociales (Kosfeld et al. 2005, Baumgartner et al. 2008, Zak et al. 2007). Kosfeld et al. (2005) mostraron, utilizando también el Juego de la confianza, que infusiones intranasales de OXT *(i)* aumentan la confianza en humanos, *(ii)* que esta acción es específica para interacciones de cooperación social, dado que *(iii)* no se explica por un aumento en el comportamiento de riesgo (*i.e.*, cuando los sujetos "confían" dinero no a otro humano, sino a una lotería), y *(iv)* que la OXT tiene un efecto específico sobre la confianza, dado que el efecto es observado solo cuando los sujetos confían, no cuando la retribuyen. Aunque los mecanismos de acción de OXT no están bien dilucidados, la evidencia sugiere que la OXT disminuye las respuestas al estrés y la ansiedad en las interacciones sociales (ver Sección 7.3). Una hipótesis plausible es que la acción principal de OXT sobre el comportamiento social humano ocurra modulando la actividad de la amígdala y de la corteza cingulada anterior, lo que dis-

minuiría el sentimiento aversivo asociado a cuando la confianza es traicionada. Un estudio de Baumgartner et al. (2008) pareciera confirmar que, efectivamente, este es el caso (ver también Englemann 2019, Fehr 2008): cuando sujetos experimentales reciben administraciones intranasales de OXT, siguen confiando en conespecíficos, a pesar de ser notificados que éstos han traicionado su confianza. Esto estaría asociado a una disminución de la actividad de la corteza cingulada anterior *antes* y de la amígdala *después* de que son notificados de que su confianza ha sido traicionada (Baumgartner et al. 2008). Una interpretación plausible es que, luego de una infusión de OXT, los sujetos están ejerciendo menos control cognitivo cuando confian en un desconocido (disminución de la actividad de corteza cingulada anterior) y que son menos sensibles a la traición en contextos sociales (disminución en la actividad de la amígdala). Por otra parte, sería esperable que el sistema de vínculos sociales estuviera involucrado en el mantenimiento de las dinámicas de confianza y reciprocidad. En esta línea, Krueger et al. (2007) mostraron que el área septal, estructura responsable de la liberación de OXT y AVP, presenta mayor actividad cuando los participantes confían, en comparación con cuando reciprocan. Este resultado es consistente con la evidencia previa que muestra que la acción OXT es específica para la confianza (Kosfeld et al. 2005).

Experimentos utilizando el Juego del dictador, por su parte, muestran que tanto las redes de mentalización como las de control cognitivo parecen estar involucradas en la manifestación de conductas altruistas: la actividad de la red de mentalización – específicamente la corteza prefrontal ventromedial y la unión temporoparietal derecha – está asociada a cuánto los sujetos valoran la ganancia material de otros (Hutcherson et al. 2015), mientras que el patrón de conectividad entre la corteza cingulada anterior y la ínsula anterior predice si una conducta altruista es impulsada por empatía

o por reciprocidad (Hein et al. 2016).

Los mecanismos descritos anteriormente son insuficientes para explicar el desarrollo de las estructuras de cooperación que observamos en las sociedades de algunos organismos. En *Homo sapiens* contemporáneo, por ejemplo, el gran tamaño de los grupos provoca que un elevado número de interacciones sean anónimas y no-repetitivas, lo que aumenta los incentivos a no cooperar (Berg et al. 1995, Coleman 2000, Camerer 2003, Nowak y Sigmund 2005) provocando, a su vez, que los niveles de confianza y reciprocidad decaigan en el transcurso del tiempo (Camerer 2003, Nowak y Sigmund 2005). Sin embargo, la presencia de mecanismos de reciprocidad indirecta, basados en la presencia de discriminadores que rehúsan cooperar con quienes no han cooperado con el resto – incluso incurriendo en un costo personal al hacerlo (Fehr y Gächter 2002, Rodríguez-Sickert et al. 2008), permiten que la cooperación en interacciones anónimas y no-repetitivas sea sostenible en el tiempo.

La reciprocidad directa puede describirse con la máxima "te hago un favor porque tú me hiciste un favor"; la reciprocidad indirecta, con la paráfrasis "te hago un favor porque algún otro me hizo un favor" (Nowak y Sigmund 2005). Modelos provenientes de la teoría de juegos evolutiva (Smith 1982) y experimentos realizados en teoría de juegos conductual (Camerer 2003) revelan las dinámicas mediante las cuales la reciprocidad indirecta sostiene (o no) la cooperación. Por ejemplo, esta no se sostiene si, dentro de una población determinada, existen individuos que siempre traicionan a quien coopera con ellos; tampoco si existen solo individuos que siempre traicionan y otros que siempre cooperan – dado que los que siempre traicionan obtendrán beneficios a expensas de los que siempre cooperan, y la estrategia de siempre traicionar terminará siendo utilizada por toda la población. Para que la cooperación sea sostenible en el tiempo, deben existir también individuos que manifiesten una

simple estrategia discriminativa, en la que siempre cooperan a menos sepan que la contraparte ha traicionado en alguna de las rondas anteriores (Brandt y Sigmund 2006, Nowak y Sigmund 2005, Imhof et al. 2005). Además, cada jugador debe llevar un rótulo ostensible, informando sobre su conducta anterior (Kandori 1992) – i.e., su reputación. Si dentro de la población existe una proporción suficiente de discriminadores, las tres estrategias coexisten durante el tiempo, impidiendo que la estrategia "siempre traicionar" se expanda a todos los individuos, y sosteniendo con esto las dinámicas de cooperación (Brandt y Sigmund 2006, Nowak y Sigmund 2005, Imhof et al. 2005).

La capacidad de discriminar en base a la conducta anterior requiere control cognitivo y mentalización (Milinski et al. 2002). Muchas veces, esta discriminación se realiza sobre la existencia de normas sociales (Kandori 1992, Coleman 2000), las que parecieran ser procesadas por la corteza prefrontal dorsomedial (Yamagishi et al. 2016) – las instituciones, por ejemplo, influencian las preferencias sociales (Rodríguez-Sickert et al. 2008). Esto permite apreciar cómo las propiedades emergentes de la interacción de agregados de individuos tienen una influencia sobre la conducta de estos sujetos: las normas sociales, por definición, no existen dentro de los individuos constituyentes del grupo, sino que son una propiedad emergente de sus interacciones. El nicho cultural de los humanos pareciera tener una influencia sobre los mecanismos de cooperación (Laland et al. 2001).

En los capítulos precedentes, hemos revisado desde las restricciones fisiológicas que permiten la gregarización, pasando por la neuroendocrinología que permite el establecimiento de lazos sociales, hasta los mecanismos neurocognitivos a la base de la manifestación de preferencias sociales, destacando cómo redes neuronales que sustentan procesos de alostasis, mentalización y control cognitivo están a la base de la

socialidad y la sociabilidad en *Homo sapiens*. En la siguiente sección, propondremos algunos organismos de la fauna local como posibles modelos para investigar estos fenómenos.

2.8 Animales nativos de Chile como potenciales modelos

Como hemos ido elaborando a lo largo del presente capítulo, existen algunos hitos evolutivos que son necesarios para comprender la evolución del comportamiento social. Uno crucial parece ser la aparición de los neuropéptidos sociales OXT y AVP, requeridos para facilitar el acercamiento prosocial y el establecimiento de lazos afectivos, principalmente entre madre y cría. Otro elemento crucial es la evolución, a partir de estos lazos, de redes neuronales relacionadas con procesos de alostasis, las que parecen ser un requisito para el desarrollo de mecanismos de neurocognición social (ver Secciones 7.5 y 7.6). Estas hipótesis son factibles de ser contrastadas en modelos de animales no-humanos. El rol de las redes de alostasis, por ejemplo, podría evaluarse mediante estudios enfocados en determinar los efectos de la deprivación social sobre las cortezas de asociación multimodal: en particular, cómo los tipos celulares de áreas donde se solapan procesos alostáticos y procesos de cognición social se ven modificadas por los distintos protocolos de deprivación.

Al menos tres organismos de la fauna nativa chilena podrían ser de interés para investigar algunos de los mecanismos neurocognitivos de sociabilidad expuestos en este capítulo: *Dromiciops gliroides*, *Cyanoliseus patagonus* y *Octodon degus*.

El monito del monte (*Dromiciops gliroides*) es un marsupial que presenta un uso comunal de refugios artificiales (Franco 2009). En este y otros mamíferos pequeños, esta conducta se ha considerado como una estrategia de termorregulación ante tem-

peraturas bajas (Ebensperger 2001). Sin embargo, en el monito del monte el uso comunal de refugios artificiales es más frecuente durante verano y otoño, estaciones caracterizadas por temperaturas ambientales altas o medias. Dado que este período coincide, además, con la actividad post-reproductiva, es más probable que los refugios comunales estén más asociados a cuidado parental que a termorregulación social (Franco 2009).

Una segunda especie potencialmente útil es el loro barranquero (*Cyanoliseus patagonus*). Si bien las cópulas extra-pareja son frecuentes en aves, incluso dentro de especies monógamas, existe una gran variabilidad entre especies en cuanto a la paternidad extra-pareja (Petrie y Kempenaers 1998, Masello et al. 2002). Utilizando marcadores moleculares en una muestra de 49 pares hembra-macho de loros barranqueros, y las crías de sus nidos, Masello et al. (2002) no encontraron marcadores de paternidad extra-pareja. Esta observación apoya que se trata de una especie que, además de social, es genéticamente monógama (*i.e.*, si bien puede presentar cópulas, no presenta paternidad extra-pareja). A la base de esto podría encontrarse el elevado nivel de cuidado paterno y el prolongado periodo de vida reproductivo del loro barranquero (Masello et al. 2002).

Finalmente, una especie que ha sido utilizada como modelo en el contexto de patologías neurodegenerativas – tales como enfermedad de Alzheimer (Ardiles et al. 2012), y en estudios en cronobiología (Ardiles et al. 2013) es *Octodon degus*, o degú. El degú es un roedor histricomorfo de desarrollo precoz, diurno, endémico de regiones semiáridas del norte y centro de Chile (Fulk 1976). Esta especie exhibe una organización social que, en algunos aspectos, ha sido comparada con la de ancestros humanos – específicamente, crías cuidadas por sus tías (Hrdy 2009, Colonnello et al. 2011). Los grupos sociales, en su estado salvaje, están conformados por 2 - 5

hembras adultas (que pueden o no estar genéticamente emparentadas; Quirici et al. 2010) y 1 - 2 machos adultos (Fulk 1976), con un índice de parentesco cercano a 0.25 (Ebensperger et al. 2004). Los grupos de degús son capaces de coordinar forrajeo tanto por contacto visual (Ebensperger y Bozinovic 2000, Colonnello et al. 2011) como por vocalizaciones (Long 2007, Colonnello et al. 2011).

En grupos sociales con dos o más hembras, estas cuidan a sus crías (generalmente 4 - 8 crías por cada hembra) en forma comunal (Ebensperger et al. 2007). Además, al contrario de ratones y ratas, los machos también contribuyen con el cuidado de las crías (Colonnello et al. 2011) – como en el caso del topillo de la pradera, mencionado en la Sección 7.3 (Young et al. 1999). La historia de vida de los degús es inusual para un roedor: nacen móviles, los ojos abiertos y el sistema auditivo ya funcional (Long y Ebensperger 2010). Presentan un periodo relativamente extendido de infancia y adolescencia, con una alta dependencia social durante su desarrollo (Colonnello et al. 2011).

Estudios conductuales muestran que la deprivación social de crías o juveniles – consistente en aislamiento total o parcial del resto de la familia – a la tercera semana de edad hace que los degús con aislamiento parcial muestren una mayor frecuencia de vocalizaciones que indican distrés (o estrés negativo), y un aumento en la búsqueda de la proximidad a familiares, en comparación con individuos control no aislados. Los degús que experimentan aislamiento total, por su parte, vocalizan menos y disminuyen su actividad locomotora, en comparación con individuos control, además de presentar alteraciones en la conducta de riesgo y neofobia (Colonnello et al. 2011). Esta misma deprivación social durante las primeras tres semanas postnatales del degú alteran el balance en redes serotoninérgicas y dopaminérgicas en la corteza prefrontal medial (Braun et al. 2000).

A pesar que en roedores las cortezas de asociación no han pasado por el dramático proceso de expansión por el que han pasado las de primates (Buckner y Krienen 2013), las características del degú lo convierten en un organismo adecuado para evaluar hipótesis de neurocognición social. Por ejemplo, es factible cuantificar la cantidad y tipos celulares de áreas cerebrales específicas en individuos sometidos a diferentes condiciones de deprivación social, como las descritas en algunos estudios previos (Braun et al. 2000, Colonnello et al. 2011). Existen reportes que muestran cambios tanto en la corteza cingulada anterior (Helmeke et al. 2001) como en la corteza infralímbica (Ovtscharoff y Braun 2001) del degú ante diferentes condiciones de deprivación. Sin embargo, hasta donde conocemos, no se han realizado estudios enfocados en evaluar los efectos de la deprivación social sobre las cortezas de asociación multimodal: en particular, cómo los tipos celulares de áreas donde se solapan procesos alostáticos y procesos de cognición social se ven modificadas por los distintos protocolos de deprivación (ver secciones 7.5 y 7.6). Por esto, la sociabilidad del degú, junto con su desarrollo precoz y su prolongado periodo de desarrollo, lo convierten en un excelente modelo de sociabilidad, en comparación con roedores altriciales.

2.9 Conclusiones

En el presente capítulo, enfocándonos en algunas especies de roedores y primates cuyos individuos forman vínculos afectivos – *i.e.*, interacciones afiliativas sociopositivas reiteradas en el tiempo –, hemos expuesto un marco que busca explicar algunos aspectos de (i) los mecanismos evolutivos y neurocognitivos que permiten la formación de grupos sociales vinculados y, sobre estos mecanismos, (ii) el desarrollo de sistemas neurocognitivos que están a la base de la expresión de dinámicas

de cooperación dentro de sistemas sociales con atributos emergentes, como las que observamos en *Homo sapiens* (Figura 1).

Primero, para manifestar una sociabilidad con vínculos se requiere que las fisiologías de los individuos no sean un impedimento para su gregarización (Seebacher y Krause 2017); a continuación, de mecanismos neurales que permitan la formación de vínculos (Young et al. 1999, Meyer-Lindenberg 2011). Los mecanismos que permiten la formación de vínculos parecieran requerir la acción de los neuropéptidos OXT y AVP. En la literatura, ambos han sido asociados con conductas que van desde la formación de lazos madre/cría hasta la manifestación de confianza, reciprocidad, altruismo y empatía (Meyer-Lindenberg 2011). Estos neuropéptidos – putativamente al inhibir la actividad de la amígdala (Englemann 2019, Baumgartner et al. 2008, Fehr 2008) – disminuyen la ansiedad y la sensibilidad a la traición en contextos sociales, facilitando el establecimiento de lazos afectivos, el reconocimiento de conespecíficos y el acercamiento prosocial.

La satisfacción de estos requerimientos mínimos para la formación de grupos con individuos vinculados entrega las condiciones necesarias para la evolución de mecanismos de cooperación complejos, en grupos de numerosos individuos, los que parecen requerir de sistemas neurocognitivos reportados abundantemente en la literatura (Hein et al. 2016, Yamagishi et al. 2016, Hutcherson et al. 2015, van den Bos et al. 2009, Chang et al. 2011, Baumgartner et al. 2008, Krueger et al. 2007, Delgado et al. 2005, King-Casas et al. 2005, Rilling et al. 2002, McCabe et al. 2001). Una parte importante de estas dinámicas de cooperación se basa en relaciones de confianza y reciprocidad (Berg et al. 1995, McCabe et al. 2003, Nowak y Sigmund 2005), las que parecen depender de la actividad de redes neuronales asociadas con procesos de mentalización (Decety y Lamm 2007) y de control cognitivo (Ullsperger et al. 2014).

Por ejemplo, el aumento de la actividad de la unión temporoparietal está asociado a conductas de altruismo (Hein et al. 2016, Hutcherson et al. 2015, van den Bos et al. 2009), mientras que la actividad de las cortezas prefrontal dorsomedial y dorsolateral pareciera ser necesaria para manifestar tanto altruismo como confianza, reciprocidad y conductas asociadas con normas sociales (Yamagishi et al. 2016, Hutcherson et al. 2015, van den Bos et al. 2009, Chang et al. 2011, Baumgartner et al. 2008, Krueger et al. 2007, Delgado et al. 2005, King-Casas et al. 2005, Rilling et al. 2002).

Estas mismas estructuras son parte de redes involucradas en procesos alostáticos (Atzil et al. 2018). Además, parecieran haber evolucionado a causa de mecanismos alométricos (Buckner y Krienen 2013). Por esto, es posible que estos mecanismos permitieran el desarrollo de las cortezas de asociación – de las que estas estructuras son parte – a partir de regiones relacionadas con procesos alostáticos. De este modo, la regulación alostática pareciera ser un proceso crucial en el desarrollo de la cognición social permitiendo, por ejemplo, la facultad de mentalización (Atzil et al., 2018, Gao et al. 2016). Estas mismas redes están sometidas a un disminuido control genético (Gómez-Robles et al. 2015) y de gradientes moleculares (Buckner y Krienen 2013), y a un aumento en la neotenia (Gao et al. 2016). Esto, junto con el alto grado de plasticidad que presentan durante la infancia, sugiere que difícilmente vienen "cableadas" de nacimiento (Atzil et al. 2018), lo que las haría más plástica a los procesos de socialización cultural y a la experiencia, sobre todo durante el desarrollo temprano.

Sin embargo, al explicar el desarrollo de un cerebro social, frecuentemente se ha puesto énfasis en una sola explicación (Dunbar y Shultz 2017) – en general, adaptacionista (Orzack y Forber 2017, Fodor 2001) –, lo que ha dificultado el establecimiento de consensos sobre la importancia relativa de los distintos mecanismos y procesos in-

volucrados en la evolución de la cognición social. Como hemos visto en este capítulo, una explicación exhaustiva debería incluir, además de presiones selectivas, (i) mecanismos alternativos como deriva génica (Ackermann y Cheverud 2004), (ii) relajo en el control genético (Gómez-Robles et al. 2015) y (iii) subproductos estructurales, como *spandrels* (Rubinov 2015), además de (iv) condiciones ecológicas – como grupos pequeños que permitieron la deriva, y (v) del desarrollo – por ejemplo, menor control sobre el destino de los progenitores celulares en las cortezas de asociación y un aumento de la neotenia en el desarrollo del cerebro, inter alia. A todo esto hay que sumar, además, la importancia de procesos de evolución cultural, los que permiten la evolución de normas sociales, haciendo plausible el desarrollo de las estructuras de cooperación sostenidas en el tiempo, mediante mecanismos de reciprocidad indirecta, como las que observamos en (al menos) las sociedades humanas (Nowak y Sigmund 2005). Como corolario, el puzle neurocognitivo del comportamiento social es bidireccional: estructuras sociales complejas han influido sobre la evolución de un cerebro social y, a su vez, la evolución de un cerebro social ha posibilitado el desarrollo de estructuras sociales complejas.

Un buen organismo modelo para investigar estos fenómenos es *Octodon degus*. Si bien se han desarrollado experimentos que utilizan su socialidad y su prolongado periodo de desarrollo para esclarecer el impacto de experiencias sociales tempranas sobre el sistema nervioso (por ejemplo, Braun et al. 2000, Colonnello et al. 2011), hasta donde los autores tienen conocimiento no se han desarrollado experimentos que busquen definir cómo estas diferentes experiencias sociales impactan sobre la cantidad y tipos celulares de áreas cerebrales relacionadas con procesos de alostasis y cognición social, en individuos sometidos a diferentes condiciones de deprivación social (ver secciones 7.5 y 7.6).

Finalmente, si bien el presente capítulo se enfocó en *Homo sapiens*, nuestra aproximación supone que la relación entre estos fenómenos observados y las redes neuronales sobre los que se implementan es bidireccional y emergente. Por esto, los autores consideran que cerebros muy distintos, en especies muy distintas, pueden ser capaces de producir fenómenos de sociabilidad y socialidad similares.

Agradecimientos

MA agradece a Luis Ebensperger (Departamento de Ecología, Pontificia Universidad Católica de Chile, Chile), Antonieta Labra (Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Noruega), y a los revisores anónimos por sus importantes comentarios y críticas en la elaboración del presente manuscrito. MA, PB y FA agradecen al artista gráfico Juan Carlos Aspé Contreras (email: juanc.aspe@gmail.com) por el diseño de las figuras de este capítulo.

2.10 Bibliografía

Aboitiz F (2018). A brain for speech. Evolutionary continuity in primate and human auditory-vocal processing. Frontiers in Neuroscience, 12, 174.

Aboitiz F (2017). A brain for speech: a view from evolutionary neuroanatomy. Springer.

Adolphs R (2009). The social brain: neural basis of social knowledge. Annual Review of Psychology. 60, 693– 716.

Ackermann RR, Cheverud JM (2004). Detecting genetic drift versus selection in human evolution. Proceedings of the National Academy of Sciences, 101(52), 17946-17951.

Althammer F, Jirikowski G, Grinevich V (2018). The oxytocin system of mice and men - similarities and discrepancies of oxytocinergic modulation in rodents and primates. Peptides. 109, 1-8.

Amodio DM, Frith CD (2006). Meeting of minds: the medial frontal cortex and social cognition. Nature Reviews Neuroscience, 7(4):268–77.

Ardiles AO, Ewer J, Acosta ML, Kirkwood A, Martínez AD, Ebensperger LA, Palacios AG (2013). Octodon degus (Molina 1782): a model in comparative biology and biomedicine. Cold Spring Harbor Protocols, 2013(4), pdb-emo071357.

Ardiles AO, Tapia-Rojas CC, Mandal M, Alexandre F, Kirkwood A, Inestrosa NC, Palacios A G (2012). Postsynaptic dysfunction is associated with spatial and object recognition memory loss in a natural model of Alzheimer's disease. Proceedings of the National Academy of Sciences, 109(34), 13835-13840.

Aspé-Sánchez M, Moreno M, Rivera MI, Rossi A, Ewer J (2016). Oxytocin and Vasopressin Receptor Gene Polymorphisms: Role in Social and Psychiatric Traits. Frontiers in Neuroscience, 9(January), 510.

Atzil S, Gao W, Fradkin I, Barrett LF (2018). Growing a social brain. Nature Human Behaviour, 2(9), 624-636.

Atzil S, Barrett LF (2017). Social regulation of allostasis: Commentary on "Mentalizing homeostasis: the social origins of interoceptive inference" by Fotopoulou & Tsakiris. Neuropsychoanalysis 19, 1–24.

Barrett LF, Satpute AB (2013). Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. Current Opinion in Neurobiology. 23, 361–372.

Baumgartner T, Heinrichs M, Vonlanthen A, Fischbacher U, Fehr E (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron, 58(4):639–50.

Beauchamp G, Fernández-Juricic E (2004). Is there a relationship between forebrain size and group size in birds? Evolutionary Ecology Research. 6, 833–842.

Beery AK (2019). Frank Beach award winner: Neuroendocrinology of group living. Hormones and Behavior, 107, 67–75.

Beery AK, Lacey EA, Francis DD (2008). Oxytocin and vasopressin receptor distributions in a solitary and a social species of tuco-tuco (Ctenomys haigi and Ctenomys sociabilis). Journal of Computational Neurology. 507, 1847–1859.

Beja-Pereira A, Luikart G, England PR, Bradley DG, Jann OC, Bertorelle G, Erhardt G (2003). Gene-culture coevolution between cattle milk protein genes and human lactase genes. Nature Genetics, 35(4), 311.

Berg J, Dickhaut J, McCabe K (1995). Trust, reciprocity, and social history. Games and Economic Behaviour, (10):122–142.

Billeke P, Soto-Icaza P, Aspé-Sánchez M, Villarroel V, Rodríguez-Sickert C (2017). Valuing others: evidence from economics, developmental psychology, and neurobiology. In: Neuroscience and Social Science (pp. 21-45). Springer, Cham.

Billeke P, Aboitiz F (2013). Social cognition in schizophrenia: from social stimuli processing to social engagement. Frontiers in Psychiatry, 4, 4.

Boyd R, Richerson PJ, Henrich J (2011). The cultural niche: Why social learning is essential for human adaptation. Proceedings of the National Academy of Sciences, 108 (Supplement 2), 10918-10925.

Brandt H, Sigmund K (2006). The good, the bad and the discriminator – errors in direct and indirect reciprocity. Journal of Theoretical Biology, 239(2), 183-194.

Braun K, Lange E, Metzger M, Poeggel G (2000). Maternal separation followed by early social deprivation affects the development of monoaminergic fiber systems in the medial prefrontal cortex of Octodon degus. Neuroscience 95, 309–318

Broad KD, Curley JP, Keverne EB (2006). Mother–infant bonding and the evolution of mammalian social relationships. Philosophical Transactions of the Royal Society B: Biological Sciences, 361(1476), 2199-2214.

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

Buckner RL, Krienen FM (2013). The evolution of distributed association networks in the human brain. Trends in Cognitive Sciences, 17(12), 648-665.

Byrne RW, Whiten A (1992). Cognitive evolution in primates: evidence from tactical deception. Man 27, 609–627.

Byrne RW, Corp N (2004) Neocortex size predicts deception rate in primates. Proceedings of the Royal Society of London B. 271, 1693–1699.

Caldwell HK, Lee HJ, Macbeth AH, Young WS (2008). Vasopressin: behavioral roles of an "original" neuropeptide. Progress in Neurobiology. 84, 1–24.

Callaerts P, Halder G, Gehring WJ (1997). PAX-6 in development and evolution. Annual Review of Neuroscience, 20(1), 483-532.

Camerer CF (2003). Behavioural Game Theory. New York. Princeton University Press.

Chang LJ, Smith A, Dufwenberg M, Sanfey AG (2011). Triangulating the neural, psychological, and economic bases of guilt aversion. Neuron, 70(3):560–72.

Churchland PS, Sejnowski T (1992). The Computational Brain (MIT Press).

Coleman J (2000). Foundations of Social Theory. Cambridge (UK). Harvard University Press.

Colonnello V, Iacobucci P, Fuchs T, Newberry RC, Panksepp J (2011). Octodon degus. A useful animal model for social-affective neuroscience research: basic description of separation distress, social attachments and play. Neuroscience & Biobehavioral Reviews, 35(9), 1854-1863.

Costa A, Torriero S, Oliveri M, Caltagirone C (2008). Prefrontal and temporoparietal involvement in taking others' perspective: TMS evidence. Behavioural Neurology, 19(1-2):71–74.

Couzin ID, Laidre ME (2009) Fission–fusion populations. Current Biology. 19, R633–R635.

Cox JC (2004). How to identify trust and reciprocity. Games and Economic Behavior, 46(2):260–281.

Conradt L, Roper TJ (2000). Synchrony and social cohesion: a fission–fusion model. Proceeding of the Royal Society of London B. 267, 2213–2218.

Clutton-Brock TH, Harvey PH (1980). Primates, brains and ecology. Journal of Zoology. 190, 309–323.

Dávid-Barrett T, Dunbar RIM (2012). Cooperation, behavioural synchrony and status in social networks. Journal of Theoretical Biology. 308, 88–95.

DeCasien AR, Williams SA, Higham JP (2017). Primate brain size is predicted by diet but not sociality. Nature Ecology & Evolution, 1(5), 0112.

Decety J, Lamm C (2007). The Role of the Right Temporoparietal Junction in Social Interaction: How Low-Level Computational Processes Contribute to Meta-Cognition. The Neuroscientist, 13(6):580–593.

Delgado MR, Frank RH, Phelps EA (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. Nature Neuroscience, 8(11):1611–8.

Donaldson ZR, Young LJ (2008). Oxytocin, vasopressin, and the neurogenetics of sociality. Science, 322(5903), 900-904.

Duffy JE, Morrison CL, Ríos RN (2000). Multiple origins of eusociality among sponge-dwelling shrimps (Synalpheus). Evolution, 54(2), 503-516.

Dunbar RIM (1998). The social brain hypothesis. Evolutionary Anthropology: Issues, News, and Reviews: 6(5), 178-190.

Dunbar RIM (2011) Constraints on the evolution of social institutions and their implications for information flow. Journal of Institutional Economics. 7, 345–371.

Dunbar RIM, Shultz S (2017). Why are there so many explanations for primate brain evolution? Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1727).

Dunbar RIM, Shultz S (2010) Bondedness and sociality. Behaviour 147, 775–803.

Dunbar RIM, Shultz S (2007). Understanding primate brain evolution. Philosophical Transactions of the Royal Society B: Biological Sciences, 362(1480), 649–658.

Ebensperger LA (2001). A review of the evolutionary causes of rodent group-living. Acta Theriologica, 46(2), 115-144.

Ebensperger LA, Hayes LD (2016). Sociobiology of caviomorph rodents: an integrative approach. John Wiley & Sons Ltd., Chichester, Reino Unido.

Ebensperger LA, Hurtado MJ, León C (2007). An experimental examination of the consequences of communal versus solitary breeding on maternal condition and the early postnatal growth and survival of degu, Octodon degus, pups. Animal Behaviour. 73, 185–194.

Ebensperger LA, Hurtado MJ, Soto-Gamboa M, Lacey EA, Chang AT (2004). Communal nesting and kinship in degus (Octodon degus). Naturwissenschaften 91, 391–395.

Ebensperger LA, Bozinovic F (2000). Communal burrowing in the hystricognath rodent, Octodon degus: a benefit of sociality? Behavioral and Ecological Sociobiology. 47, 365–369.

Ebstein RP, Knafo A, Mankuta D, Chew SH, Lai PS (2012). The contributions of oxytocin and vasopressin pathway genes to human behavior. Hormones and Behaviour. 61, 359–379.

Engelmann JB, Meyer F, Ruff CC, Fehr E (2019). The neural circuitry of affect-induced distortions of trust. Science Advances, 5(3). Falótico T, Proffitt T, Ottoni EB, Staff RA, Haslam M (2019). Three thousand years of wild capuchin stone tool use. Nature Ecology & Evolution, 1.

Farine DR, Montiglio PO, Spiegel O (2015) From individuals to groups and back: the evolutionary implications of group phenotypic composition. Trends in Ecology

and Evolution. 30, 609–621.

Fehr E (2008). The effect of neuropeptides on human trust and altruism: a neuroeconomic perspective. In Hormones and Social Behaviour (pp. 47-56). Springer, Berlin, Heidelberg.

Fehr E, Fischbacher U (2003). The nature of human altruism. Nature, 425(6960), 785.

Fehr E, Fischbacher U (2002). Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. The Economic Journal, 112(478), C1-C33.

Fehr E, Gächter S (2002). Altruistic punishment in humans. Nature, 415(6868), 137-140.

Fehr E, Schmidt KM (2001). Theories of fairness and reciprocity-evidence and economic applications. CEPR Discussion Paper, No. 2703.

Feldman MW, Laland KN (1996). Gene-culture coevolutionary theory. Trends in Ecology & Evolution, 11(11), 453-457.

Felleman DJ, van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex 1, 1–47

Ferretti V, Maltese F, Contarini G, Nigro M, Bonavia A, Huang H, Castellani G (2019). Oxytocin signaling in the central amygdala modulates emotion discrimination in mice. Current Biology, 29(12), 1938-1953.

Fink S, Excoffier L, Heckel G (2006). Mammalian monogamy is not controlled by a single gene. Proceedings of the National Academy of Sciences U.S.A. 103, 10956–10960.

Fodor JA (2001). The mind doesn't work that way: The scope and limits of computational psychology. MIT press.

Fodor JA (1986). La modularidad de la mente: un ensayo sobre la psicología de las facultades. Ediciones Morata.

Forsythe R, Horowitz JL, Savin NE, Sefton, M (1994). Fairness in simple bargaining experiments. Games and Economic Behaviour,(6):347–369.

Fox KCR, Muthukrishna M, Shultz S (2017). The social and cultural roots of whale and dolphin brains. Nature Ecology and Evolution, 1(11), 1699–1705.

Fox MD, Corbetta M, Snyder AZ, Vincent JL, Raichle ME (2006). Spontaneous neu-

ronal activity distinguishes human dorsal and ventral attention systems. Proceedings of the National Academy of Sciences, 103(26), 10046-10051.

Franco LM (2009). Importancia del agrupamiento durante el sopor: mecanismos proximales y consecuencias ecológicas en el marsupial ancestral, monito del monte (Dromiciops gliroides, MARSUPIALIA: MICROBIOTHERIA). Ph.D. tesis, Universidad Austral de Chile, Valdivia, Chile.

Frith CD, Frith U (1999). Interacting Minds - A Biological Basis. Science, 286(5445):1692–695.

Fulk GW (1976). Notes on the activity, reproduction, and social behavior of Octodon degus. Journal of Mammalogy. 57, 495–505.

Furman DJ, Chen MC, Gotlib IH (2011). Variant in oxytocin receptor gene is associated with amygdala volume. Psychoneuroendocrinology 36, 891–897.

Fuster JM (2000). The module: crisis of a paradigm. Neuron, 26(1), 51-53.

Gadagkar R (1987). What are Social Insects? IUSSI Indian Chapter Newsletter, 1(2), 3-4.

Gao W, Gilmore JH, Shen D, Smith JK, Zhu H, Lin W (2013). The synchronization within and interaction between the default and dorsal attention networks in early infancy. Cerebral Cortex, 23(3), 594-603.

Gao W, Alcauter S, Elton A, Hernandez-Castillo CR, Smith JK, Ramirez J, Lin W (2015). Functional network development during the first year: relative sequence and socioeconomic correlations. Cerebral Cortex, 25(9), 2919-2928.

Gao W, Lin W, Grewen K, Gilmore JH (2016). Functional connectivity of the infant human brain plastic and modifiable. Neuroscientist 23, 169–184.

Garrison JL, Macosko EZ, Bernstein S, Pokala N, Albrecht DR, Bargmann CI (2012). Oxytocin/vasopressin-related peptides have an ancient role in reproductive behavior. Science 338, 540–543.

Goldman-Rakic PS (1988) Topography of cognition: parallel distributed networks in primate association cortex. Annual Review in Neuroscience. 11, 137–156

Gómez-Robles A, Hopkins WD, Schapiro SJ, Sherwood CC (2015). Relaxed genetic control of cortical organization in human brains compared with chimpanzees. Proceedings of the National Academy of Sciences, 112(48), 14799-14804.

González-Forero M, Gardner A (2018). Inference of ecological and social drivers of human brain-size evolution. Nature, 557(7706), 554.

Gould SJ, Eldredge N (1972). Punctuated equilibria: an alternative to phyletic gradualism. Essential Readings in Evolutionary Biology, 82-115.

Gould SJ, Lewontin RC (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proceedings of the Royal Society of London. Series B. Biological Sciences, 205(1161), 581-598.

Gould SJ, Vrba ES (1982). Exaptation: a missing term in the science of form. Paleobiology, 8(1), 4-15.

Gowlett J, Gamble C, Dunbar RIM (2012) Human evolution and the archaeology of the social brain. Current Anthropology. 53, 693–722.

Hamilton WD (1996). Narrow Roads of Gene Land Vol. 1. New York. Freeman.

Hammock EAD, Lim MM, Nair HP, Young LJ (2005). Association of vasopressin 1a receptor levels with a regulatory microsatellite and behavior. Genes, Brain and Behaviour. 4, 289–301.

Hammock EAD, Young LJ (2005). Microsatellite instability generates diversity in brain and sociobehavioral traits. Science 308, 1630–1634.

Hammock EAD, Young LJ (2004). Functional microsatellite polymorphism associated with divergent social structure in vole species. Molecular Biology and Evolution. 21, 1057–1063.

Hanson IM, Fletcher JM, Jordan T, Brown A, Taylor D, Adams RJ, van Heyningen V (1994). Mutations at the PAX6 locus are found in heterogeneous anterior segment malformations including Peters' anomaly. Nature Genetics, 6(2), 168.

Haug H (1987). Brain sizes, surfaces, and neuronal sizes of the cortex cerebri: a stereological investigation of man and his variability and a comparison with some mammals (primates, whales, marsupials, insectivores, and one elephant). American Journal of Anatomy, 180(2), 126-142.

Healy SD, Rowe C (2007) A critique of comparative studies of brain size. Proceedings of the Royal Society B, 274, 453–464.

Henrich JP, Boyd R, Bowles S, Fehr E, Camerer C, Gintis H (Eds.). (2004). Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies. Oxford University Press on Demand.

Hein G, Morishima Y, Leiberg S, Sul S, Fehr E (2016). The brain's functional network architecture reveals human motives. Science, 351(6277):1074–1078.

Helmeke C, Poeggel G, Braun K (2001). Differential emotional experience induces

elevated spine densities on basal dendrites of pyramidal neurons in the anterior cingulate cortex of Octodon degus. Neuroscience, 104(4), 927-931.

Herbert-Read JE, Perna A, Mann RP, Schaerf TM, Sumpter DJ, Ward AJ (2011) Inferring the rules of interaction of shoaling fish. Proceedings of the National Academy of Sciences USA, 108, 18 726–18 731.

Herculano-Houzel S, Messeder DJ, Fonseca-Azevedo K, Pantoja NA (2015). When larger brains do not have more neurons: increased numbers of cells are compensated by decreased average cell size across mouse individuals. Frontiers in Neuroanatomy, 9, 64.

Holroyd CB, Coles MG (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychological Review, 109(4), 679.

Hrdy SB (2009). Mothers and Others. The Belknap Press of Harvard University Press, Cambridge, MA.

Huber D, Veinante P, Stoop R (2005). Vasopressin and oxytocin excite distinct neuronal populations in the central amygdala. Science, 308(5719), 245-248.

Hutcherson CA, Bushong B, Rangel A (2015). A Neurocomputational Model of Altruistic Choice and Its Implications. Neuron, 87(2):451–463.

Imhof L, Fudenberg D, Nowak MA (2005). Evolutionary cycles of cooperation and defection. Proceedings of the. National Academy of Sciences USA, 102, 10797–10800.

Inoue H, Yamasue H, Tochigi M, Takei K, Suga M, Abe O (2010). Effect of tryptophan hydroxylase-2 gene variants on amygdalar and hippocampal volumes. Brain Research, 1331, 51–57.

Insel TR, Shapiro LE (1992). Oxytocin receptor distribution reflects social organization in monogamous and polygamous voles. Proceedings of the National Academy of Sciences USA. 89, 5981–5985.

Insel TR (2010). The challenge of translation in social neuroscience: a review of oxytocin, vasopressin, and affiliative behavior. Neuron, 65, 768–779.

Johnson ND, Mislin AA (2011). Trust games: A meta-analysis. Journal of Economic Psychology, 32(5), 865-889.

Jouventin P, Christen Y, Dobson FS (2016). Altruism in wolves explains the coevolution of dogs and humans. Ideas in Ecology and Evolution, 9(1).

Kano F, Krupenye C, Hirata S, Tomonaga M, Call J (2019). Great apes use self-

experience to anticipate an agent's action in a false-belief test. Proceedings of the National Academy of Sciences, 116(42), 20904-20909.

Kandori, M (1992). Social norms and community enforcement. Reviews in Economics Studies, 59, 63–80.

Kappeler PM, van Schaik CP (2002). Evolution of primate social systems. International Journal of Primatology, 23:707-740.

King C, Stanfield WD (1997). Panmixia (panmixis): random mating as contrasted with assortative mating. Dictionary of Genetics. Oxford.

King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR, Montague PR (2005). Getting to know you: reputation and trust in a two-person economic exchange. Science, (New York, N.Y.), 308(5718):78–83.

Kleckner IR, Zhang J, Touroutoglou A, Chanes L, Xia C, Simmons WK, Barrett LF (2017). Evidence for a large-scale brain system supporting allostasis and interoception in humans. Nature Human Behaviour, 1(5), 0069.

Knafo A, Israel S, Darvasi A, Bachner-Melman R, Uzefovsky F, Cohen L (2008). Individual differences in allocation of funds in the dictator game associated with length of the arginine vasopressin 1a receptor RS3 promoter region and correlation between RS3 length and hippocampal mRNA. Genes, Brain and Behaviour. 7, 266–275.

Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E (2005). Oxytocin increases trust in humans. Nature, 435, 673–676.

Krause J, Lusseau D, James R (2009). Animal social networks: an introduction. Behavioral and Ecological Sociobiology. 63, 967–973.

Krause J, Ruxton GD (2002). Living in groups. Oxford University Press, Oxford, Reino Unido.

Krueger F, Parasuraman R, Iyengar V, Thornburg M, Weel J, Lin M (2012). Oxytocin receptor genetic variation promotes human trust behavior. Frontiers in Human Neurosciences. 6:4.

Krueger F, McCabe K, Moll J, Kriegeskorte N, Zahn R, Strenziok M, Heinecke A, Grafman J (2007). Neural correlates of trust. Proceedings of the National Academy of Sciences of the United States of America, 104(50):20084–9.

Kurvers RHJ, Krause J, Croft DP, Wilson ADM, Wolf M (2014) The evolutionary and ecological consequences of animal social networks: emerging issues. Trends in Ecology and Evolution. 29, 326–335.

Lacey EA, Sherman PW (2007). The ecology of sociality in rodents. In: Wolff, J.O., Sherman, P.W. (Eds.), Rodent Societies: An Ecological & Evolutionary Perspective. University of Chicago Press, Chicago, pp. 243–254.

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. Nature, 413(6855), 519.

Laland K, Odlee-Smee J, Feldman M (2000). Niche construction, biological evolution, and cultural change. Behavioral and Brain Sciences, 23(1), 131-46; discussion 146-75.

Laland KN, Odling-Smee J, Feldman MW (2001). Cultural niche construction and human evolution. Journal of Evolutionary Biology, 14(1), 22-33.

Laland KN, O'Brien MJ (2011). Cultural niche construction: An introduction. Biological Theory, 6(3), 191-202.

Lamm C, Batson CD, Decety J (2007). The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. Journal of Cognitive Neuroscience, 19(1), 42-58.

Landgraf R, Neumann ID (2004). Vasopressin and oxytocin release within the brain: a dynamic concept of multiple and variable modes of neuropeptide communication. Frontiers in Neuroendocrinology. 25, 150–176.

Lee PC (1994). Social structure and evolution. In: Slater, P.J.B., Halliday, T. (Eds.), Behaviour and Evolution. Cambridge University Press, New York, NY, USA, pp. 266–303.

Leng G, Ludwig M (2008). Neurotransmitters and peptides: whispered secrets and public announcements. Journal of Physiology. 586(23), 5625–5632.

Litingtung Y, Chiang C (2000). Control of Shh activity and signaling in the neural tube. Developmental dynamics: an official publication of the American Association of Anatomists, 219(2), 143-154.

Long CV (2007). Vocalisations of the degu Octodon degus, a social caviomorph rodent. Bioacoustics 16, 223–244.

Long CV, Ebensperger LA (2010). Pup growth rates and breeding female weight changes in two populations of captive bred degus (Octodon degus), a precocial caviomorph rodent. Reproduction in Domestic Animals, 45(6), 975-982.

Loup F, Tribollet E, Dubois-Dauphin M, Dreifuss JJ (1991). Localization of high-

affinity binding sites for oxytocin and vasopressin in the human brain. An autoradiographic study. Brain Research. 555, 220–232.

Luo S, Li B, Ma Y, Zhang W, Rao Y, Han S (2015). Oxytocin receptor gene and racial ingroup bias in empathy-related brain activity. Neuroimage 110, 22–31.

MacDonald K, MacDonald TM (2010). The peptide that binds: a systematic review of oxytocin and its prosocial effects in humans. Harvard Review of Psychiatry 18, 1–21.

MacLean EL, Hare B, Nunn CL, Addessi E, Amici F, Anderson RC, Boogert NJ (2014). The evolution of self-control. Proceedings of the National Academy of Sciences, 111(20), E2140-E2148.

Margulies DS, Ghosh SS, Goulas A, Falkiewicz M, Huntenburg JM, Langs G (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. Proceedings of the National Academy of Sciences. 113, 12574–12579.

Masello JF, Sramkova A, Quillfeldt P, Epplen JT, Lubjuhn T (2002). Genetic monogamy in burrowing parrots Cyanoliseus patagonus?. Journal of Avian Biology, 33(1), 99-103.

McCabe KA, Rigdon ML, Smith VL (2003). Positive reciprocity and intentions in trust games. Journal of Economic Behavior and Organization, 52(2):267–275.

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 of the United States of America, 98(20):11832–5.

Meyer-Lindenberg A, Domes G, Kirsch P, Heinrichs M (2011). Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. Nature Review Neuroscience, 12, 524–538.

Mesulam MM (1990) Large-scale neurocognitive networks and distributed processing for attention, language, and memory. Annual Review in Neurology, 28, 597–613.

Milinski M, Semmann D, Krambeck HJ (2002). Reputation helps solve the 'tragedy of the commons'. Nature, 415, 424–426.

Miller GF (1999). Sexual selection for cultural displays. In The evolution of culture (eds RIM Dunbar, C Knight, C Power), pp. 71–91. Edinburgh, UK: Edinburgh University Press.

Millstein RL (2017). Genetic Drift. The Stanford Encyclopedia of Philosophy (Fall 2017 Edition). Edward N. Zalta (ed.)

Moll H, Tomasello M (2007). Cooperation and human cognition: the Vygotskian intelligence hypothesis. Philosophical Transactions of the Royal Society B, 362, 639–648.

Morrison RE, Groenenberg M, Breuer T, Manguette ML, Walsh PD (2019). Hierarchical social modularity in gorillas. Proceedings of the Royal Society B, 286(1906), 20190681.

Murphy D (2003). Adaptationism and psychological explanation. In Evolutionary Psychology (pp. 161-184). Springer, Boston, MA.

Nakajima M, Görlich A, Heintz N (2014). Oxytocin modulates female sociosexual behavior through a specific class of prefrontal cortical interneurons. Cell, 159, 295–305.

Navarrete AF, Reader SM, Street SE, Whalen A, Laland KN (2016). The coevolution of innovation and technical intelligence in primates. Philosophical Transactions of the Royal Society B, 371, 20150186.

Nithianantharajah J, Komiyama NH, McKechanie A, Johnstone M, Blackwood DH, St Clair D, Grant SG (2013). Synaptic scaffold evolution generated components of vertebrate cognitive complexity. Nature Neuroscience, 16(1), 16.

Nowak MA, Sigmund K (2005). Evolution of indirect reciprocity. Nature, 437(October), 1291–11298. Ord TJ, Martins EP (2010). The evolution of behavior: phylogeny and the origin of present-day diversity. Evolutionary behavioral ecology, 108-128.

Orzack SH, Forber P(2017). "Adaptationism", The Stanford Encyclopedia of Philosophy, (Spring 2017 Edition), Edward N. Zalta (ed.)

Ostner J, Schülke O (2018). Linking sociality to fitness in primates: a call for mechanisms. Advances in the Study of Behavior, 50:127-175.

Ovtscharoff Jr W, Braun K (2001). Maternal separation and social isolation modulate the postnatal development of synaptic composition in the infralimbic cortex of Octodon degus. Neuroscience, 104(1), 33-40.

Petrie M, Kempenaers B (1998). Extra-pair paternity in birds: explaining variation between species and populations. Trends in Ecology and Evolution. 13: 52 – 58.

Pitnick S, Jones KE, Wilkinson GS (2006). Mating system and brain size in bats. Proceedings of the Royal Society B, 273, 719–724.

Posner MI, Snyder CRR (1975). Attention and cognitive control. In R. Solso (ed.), Information Processing and Cognition: The Loyola Symposium. Hillsdale, N.J.:

Lawrence Erlbaum Associates.

Powell LE, Isler K, y Barton RA (2017). Re-evaluating the link between brain size and behavioural ecology in primates. Proceedings of the Royal Society B: Biological Sciences, 284(1865), 20171765.

Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Petersen SE (2011). Functional network organization of the human brain. Neuron, 72(4), 665-678.

Premack D, Woodruff D (1978). Does the chimpanzee have a 'theory of mind'? Behavioral Brain Sciences. 4, 515–526. Quirici V, Faugeron S, Hayes LD, Ebensperger LA (2010). Absence of kin structure in a population of the group-living rodent Octodon degus. Behavioral Ecology, 22(2), 248-254.

Raichle ME (2015). The brain's default mode network. Annual Review in Neuroscience 38:433–447.

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL (2001). A default mode of brain function. Proceedings of the National Academy of Sciences, 98(2), 676-682.

Ramírez-Barrantes R, Arancibia M, Stojanova J, Aspé-Sánchez M, Córdova C, Henríquez-Ch RA (2019). Default Mode Network, Meditation, and Age-Associated Brain Changes: What Can We Learn from the Impact of Mental Training on Well-Being as a Psychotherapeutic Approach?. Neural Plasticity, 2019.

Reader SM, Laland KN (2002). Social intelligence, innovation, and enhanced brain size in primates. Proceedings of the National Academy of Sciences, 99, 4436–4441.

Reader SM, Hager Y, Laland KN (2011). The evolution of primate general and cultural intelligence. Philosophical Transactions of the Royal Society B, 366, 1017–1027.

Réale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ (2007). Integrating animal temperament within ecology and evolution. Biological Reviews 82:291-318.

Rilling J, Gutman D, Zeh T, Pagnoni G, Berns G, Kilts C (2002). A neural basis for social cooperation. Neuron, 35(2):395–405.

Rodrigues SM, Saslow LR, Garcia N, John OP, Keltner D (2009). Oxytocin receptor genetic variation relates to empathy and stress reactivity in humans. Proceedings of the National Academy of Sciences, 106, 21437–21441.

Rodriguez-Sickert C, Guzmán RA, Cárdenas JC (2008). Institutions influence preferences: Evidence from a common pool resource experiment. Journal of Economic

Behavior & Organization, 67(1), 215-227.

Rosati AG (2017). Foraging cognition: reviving the ecological intelligence hypothesis. Trends in Cognitive Sciences, 21, 691–702 (2017).

Rubinov, M. (2016). Constraints and spandrels of interareal connectomes. Nature communications, 7(1), 1-11.

Sachser N, Dürschlag M, Hirzel D (1998). Social relationships and the management of stress. Psychoneuroendocrinology 23:891-904.

Sadino JM, Donaldson ZR (2018). Prairie voles as a model for understanding the genetic and epigenetic regulation of attachment behaviors. ACS Chemical Neuroscience, 9(8), 1939-1950.

Sala M, Braida D, Lentini D, Busnelli M, Bulgheroni E, Capurro V (2015). Pharmacologic rescue of impaired cognitive flexibility, social deficits, increased aggression, and seizure susceptibility in oxytocin receptor null mice: a neurobehavioral model of autism. Biological Psychiatry, 69, 875–882.

Samson D, Apperly IA, Chiavarino C, Humphreys GW (2004). Left temporoparietal junction is necessary for representing someone else's belief. Nature Neuroscience, 7(5), 499-500.

Schillaci MA, Grant SGN (2006). Sexual selection and the evolution of brain size in primates. PLoS ONE, 1, e62.

Schroeder L, Ackermann RR (2017). Evolutionary processes shaping diversity across the Homo lineage. Journal of Human Evolution, 111, 1-17.

Seebacher F, Krause J (2017). Physiological mechanisms underlying animal social behaviour. Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1727).

Selemon LD, Goldman-Rakic PS (1988). Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: evidence for a distributed neural network subserving spatially guided behavior. Journal of Neuroscience 8, 4049–4068

Shenhav A, Botvinick MM, Cohen JD (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron, 79(2):217–40.

Sheppard CE, Marshall HH, Inger R, Thompson FJ, Vitikainen EI, Barker S, Cant MA (2018). Decoupling of genetic and cultural inheritance in a wild mammal. Current Biology, 28(11), 1846-1850.

Shingleton A (2010). Allometry: The Study of Biological Scaling. Nature Education Knowledge 3(10):2.

Shultz S, Nelson E, Dunbar RIM (2012). Hominin cognitive evolution: identifying patterns and processes in the fossil and archaeological record. Philosophical Transactions of the Royal Society B, 367, 2130–2140.

Shultz S, Opie C, Atkinson QD (2011). Stepwise evolution of stable sociality in primates. Nature, 479(7372), 219–222.

Shultz S, Dunbar RIM (2007). The evolution of the social brain: anthropoid primates contrast with other vertebrates. Proceedings of the Royal Society B, 274, 2429–2436.

Smaers J, Soligo C (2013). Brain reorganization, not relative brain size, primarily characterizes anthropoid brain evolution. Proceedings of the Royal Society B: Biological Sciences.

Smith JM (1982). Evolution and the Theory of Games. Cambridge University Press.

Smith K (2002). The cultural evolution of communication in a population of neural networks. Connection Science, 14(1), 65–84.

Sober E (2014). The nature of selection: Evolutionary theory in philosophical focus. University of Chicago Press.

Soto-Icaza P, Vargas L, Aboitiz F, Billeke P (2019). Beta oscillations precede joint attention and correlate with mentalization in typical development and autism. Cortex, 113, 210-228.

Spunt RP, Adolphs R. (2017). A new look at domain specificity: insights from social neuroscience. Nature Reviews Neuroscience, 18(9), 559.

Sterling P (2012). Allostasis: a model of predictive regulation. Physiology and Behaviour. 106, 5–15.

Stiglitz JE, Walsh CE (1993). Economics. New York/London: W.

Suzuki IK, Gacquer D, Van Heurck R, Kumar D, Wojno M, Bilheu A, Detours V (2018). Human-specific NOTCH2NL genes expand cortical neurogenesis through Delta/Notch regulation. Cell, 173(6), 1370-1384.

Tansey KE, Hill MJ, Cochrane LE, Gill M, Anney RJ, Gallagher L (2011). Functionality of promoter microsatellites of arginine vasopressin receptor 1A (AVPR1A): implications for autism. Molecular Autism, 2, 3.

Terborgh J, Janson CH (1986). The socioecology of primate groups. Annual Review

of Ecology and Systematics, 17(1), 111-136.

Thompson E, Varela FJ (2001). Radical embodiment: neural dynamics and consciousness. Trends in Cognitive Sciences, 5(10), 418-425.

Trivers R (1985). Social Evolution. Benjamin/Cummings. Reading, MA.

Ullsperger M, Fischer AG, Nigbur R, Endrass T (2014). Neural mechanisms and temporal dynamics of performance monitoring. Trends in Cognitive Sciences, 18(5):259–267.

van den Bos W, van Dijk E, Westenberg M, Rombouts SRB, Crone, E (2009). What motivates repayment? Neural correlates of reciprocity in the Trust Game. Social Cognitive and Affective Neuroscience, 4(3):294–304.

van Schaik CP, Isler K, Burkart JM (2012). Explaining brain size variation: from social to cultural brain. Trends in Cognitive Sciences, 16, 277–284.

Vogeley K (2017). Two social brains: Neural mechanisms of intersubjectivity. Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1727), 1–11.

Weaver TD, Roseman CC, Stringer CB (2007). Were Neanderthal and modern human cranial differences produced by natural selection or genetic drift? Journal of Human Evolution, 53(2), 135-145.

Weintraub ER (2007). Neoclassical Economics. En The Concise Encyclopedia of Economics.

Whiten A, Byrne RW. (1997). Machiavellian Intelligence II: Extensions and Evaluations. Cambridge Univ. Press, Cambridge, UK.

Wilkinson GS (1988). Reciprocal altruism in bats and other mammals. Ethology and Sociobiology, 9(2-4), 85-100.

Wilson EO (1971). The insect societies. No. QL 496. W54.

Wilson EO (1975). Sociobiology. Cambridge. Harvard Univ. Press.

Yamagishi T, Takagishi H, Fermin ASR, Kanai R, Li Y, Matsumoto Y (2016). Cortical thickness of the dorsolateral prefrontal cortex predicts strategic choices in economic games. Proceedings of the National Academy of Sciences, 113(20), 5582–5587.

Young L, Nilsen R, Waymire K, MacGregor G (1999). Increased affillative response to vasopressin in mice expressing the V1a receptor from a monogamous vole. Nature, 400, 1998–2000.

Yuste R (2011). Dendritic spines and distributed circuits. Neuron 71, 772–781.

Yuste R, Tank DW (1996). Dendritic integration in mammalian neurons, a century after Cajal. Neuron, 16(4), 701-716.

Zak PJ, Stanton AA, Ahmadi S. (2007). Oxytocin increases generosity in humans. PLoS ONE 2:e1128.

Zink CF, Meyer-Lindenberg A (2012). Human neuroimaging of oxytocin and vasopressin in social cognition. Hormones and Behaviour. 61, 400–409.

3 Oxytocin and vasopressin receptor gene polymorphisms: role in social and psychiatric traits

Paper published in *Frontiers in Neurosciences* (2016, Vol. 9, pp. 510) by:

Aspé-Sánchez, M.[1,3,5], Moreno, M.[2], Rivera, M. I.[3], Rossi, A.[4] & Ewer, J.[1]

[1] Laboratorio de Neurogenética y Desarrollo, Centro Interdisciplinario de Neurociencia de Valparaíso (CINV), Universidad de Valparaíso, Valparaíso, Chile.

[2] Cognitive Neurosciences Laboratory, Department of Psychiatry, Pontificia Universidad Católica de Chile, Santiago, Chile.

[3] Centro de Investigación en Complejidad Social (CICS), Facultad de Gobierno, Universidad del Desarrollo, Santiago, Chile.

[4] Cognitive Science Program, Indiana University, Bloomington, IN. USA.

[5] Cognitive Neuroscience Sector, Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy.

[*] Correspondence: Mauricio Aspé-Sánchez, Centro Interdisciplinario de Neurociencia de Valparaíso (CINV), Laboratorio de Neurogenética y Conducta, Gran Bretaña 1111, Playa Ancha, Valparaíso, Chile. mauricio.aspe.s@gmail.com

Keywords: SNP1, SSR2, psychiatric disorders3, dopamine4, serotonin5, polygenic trait6, quantitative traits loci7, GWAS8.

3.1 Abstract

Oxytocin (OXT) and arginine-vasopressin (AVP) are two phylogenetically conserved neuropeptides that have been implicated in a wide range of social behaviors. Although a large body of research, ranging from rodents to humans, has reported on the effects of OXT and AVP administration on affiliative and trust behaviors, and has highlighted the genetic contributions of OXT and AVP receptor polymorphisms to both social behaviors and to diseases related to social deficits, the consequences of peptide administration on psychiatric symptoms, and the impact of receptor polymorphisms on receptor function, are still unclear. Despite the exciting advances that these reports have brought to social neuroscience, they remain preliminary and suffer from the problems that are inherent to monogenetic linkage and association studies. As an alternative, some studies are using polygenic approaches, and consider the contributions of other genes and pathways, including those involving DA, 5-HT, and reelin, in addition to OXT and AVP; a handful of report are also using genome-wide association studies.

This review summarizes findings on the associations between OXT and AVP receptor polymorphism, social behavior, and psychiatric diseases. In addition, we discuss reports on the interactions of OXT and AVP receptor genes and genes involved in other pathways (like those of dopamine, serotonin, and reelin), as well as research that has shed some light on the impact of gene polymorphisms on the volume, connectivity, and activation of specific neural structures, differential receptor expression, and plasma levels of the OXT and AVP peptides. We hope that this effort will be helpful for understanding the studies performed so far, and for encouraging the inclusion of other candidate genes not explored to date.

3.2 Introduction

Oxytocin (OXT) and arginine-vasopressin (AVP) are, nowadays, perhaps the most interesting molecules for social neuroscience (Insel, 2010; Meyer-Lindenberg et al., 2011; Zink et al., 2012). They are two closely related, phylogenetically conserved nonapeptides, which originated more than 700 million years ago (MacDonald and MacDonald, 2010) and differ in only two aminoacids (Caldwell et al., 2008; McDonald and McDonald, 2010; Insel, 2010): whereas the aminoacid sequence of OXT includes an isoleucine and a leucine at the third and eighth position, respectively, AVP has a phenylalalanine and an arginine in the corresponding positions. Both peptides contain two cysteine residues that form a disulfide bond, creating a cyclic six aminoacid ring (Caldwell et al., 2008).

OXT and AVP are synthesized in the somas of two groups of neurons located in the paraventricular and supraoptic nucleus of the hypothalamus; they are then transported along axons that project to the posterior pituitary, from where they are ultimately released into the circulation, acting as hormones on peripheral targets (Loup et al., 1991; Meyer-Lindenberg et al., 2011; Zink et al., 2012). In addition to their well-documented peripheral actions (e.g., that of OXT in uterine contraction and milk ejection, and of AVP in antidiuresis and vasoconstriction; Loup et al., 1991) they also function within the brain as neuromodulators or neurotransmitters. There they act mainly via dendritic release (Landgraf and Neumann, 2004; Leng and Ludwig, 2008; Ludwig and Leng, 2006; Neumann and Landgraf, 2012; Stoop, 2012) to regulate complex social cognition and behaviors, including attachment (Insel, 2010; Young et al., 1999) and social exploration and recognition (Meyer-Lindenberg, 2011). Furthermore, there is a growing body of evidence showing that these neuropeptide signaling pathways are impaired in mental disorders associated with social deficits

(Meyer-Lindenberg, 2011).

This review summarizes findings on the impact of different neuropeptide receptor polymorphism on social cognition and behaviors, and some of the most common mental disorders associated with deficits in social function. Our goal is to put special emphasis on the studies that have shed light on associations involving different receptor polymorphism, even if the functional significance of these polymorphisms is currently unknown. In addition, we highlight the importance of polygenic approaches for a fruitful understanding of the OXT and AVP signaling mechanisms, and highlight studies on the effect of intranasal OXT and AVP infusions on psychiatric symptoms.

3.2.1 OXT and AVP receptors and their polymorphisms

OXT and AVP receptors (OXTR and AVPR, respectively) belong to the seven transmembrane domain G-protein coupled receptor super-family. Whereas only one form of OXTR is known, three subtypes of AVPR exists, called AVPR1a, AVPR1b, and AVPR2 (Ebstein et al., 2012). The AVPR1a and AVPR1b subtypes are both coupled to Gq, and signal via phospholipase C. The inositol triphosphate (IP3) produced upon receptor activation induces the mobilization of calcium (Ca2+) from intracellular reservoirs; it also causes the release of diacylglycerol (DAG) which, in addition to increasing intracellular Ca2+, also activates protein kinase C (PKC). The rise in intracellular calcium levels activates map kinase (MAPK) and calcium/calmodulin-dependent protein kinase II (CAMKII). In addition, activation of this signaling pathway produces a depolarization that leads to an extracellular Ca2+ entry via voltage-gated calcium channels (Thibonnier et al., 2000; 1995; 1992). AVPR1a and AVPR1b are expressed in several tissue and organs, including platelets, the adrenal

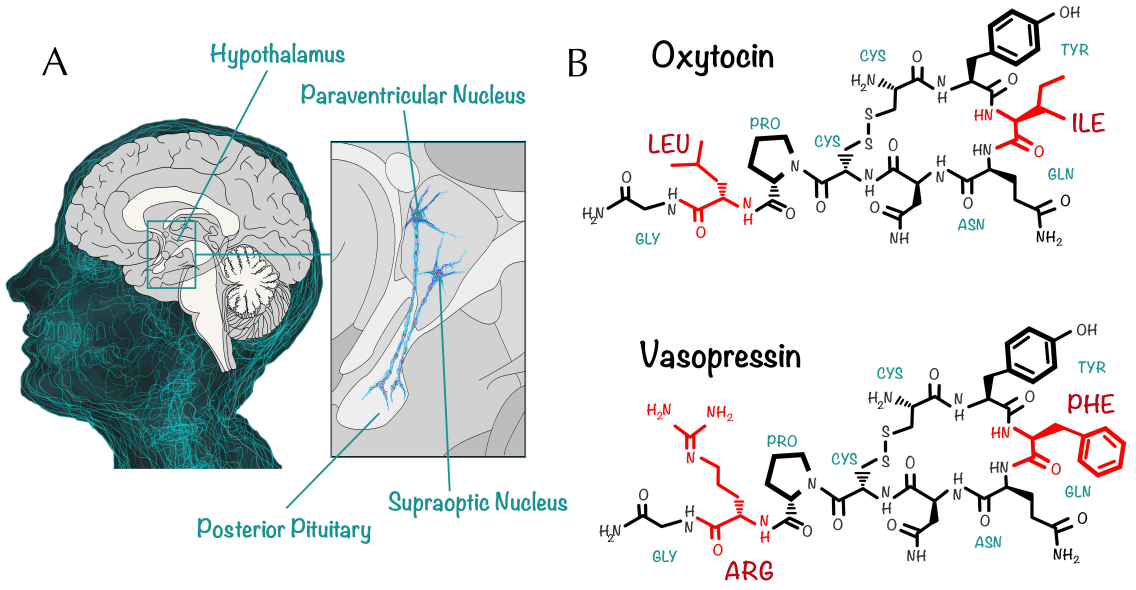


Figure 5: OXT and AVP are two closely related nonapeptides that exert their action on central and peripheral targets. **(A)** OXT and AVP are synthesized in the PVN and the SON of the hypothalamus. The peptidergic neurons in these nuclei project axons to the posterior pituitary, from where the peptides are released into the circulation. They act as hormones on peripheral targets, having well-documented actions (uterine contraction and vasoconstriction, for instance). In addition, dendrites of neurons in the PVN and the SON release the peptides directly into the brain, where they act as neurotransmitters or neuromodulators, regulating complex social cognition and behaviors. **(B)** OXT and AVP differ in only two aminoacids: this schematic drawing shows that, whereas the aminoacid sequence of OXT (top) includes an isoleucine at the third and a leucine at the eighth position, that of AVP (bottom) includes a phenylalanine and an arginine in the corresponding positions. Both peptides contain a cyclic six aminoacid ring, because of the disulfide bond formed by two cysteine residues.

cortex, kidney, spleen, smooth muscle, endothelium, and adipocytes (Thibonnier et al., 2000; 1995; 1992). In the human brain, they are expressed in the lateral septum, thalamus, basal amygdaloid nucleus, and brainstem but, interestingly, not in the cortex (Meyer-Lindenberg et al., 2009). The AVPR2 subtype, on the other hand, is coupled to Gs, which activates adenylyl cyclase (AC) and causes the production

of cAMP and the activation of protein kinase A (PKA). This receptor subtype is expressed on the basolateral membrane of the colleting duct in the medullary portion of the kidney (Thibonnier et al., 2000).

The single OXTR is coupled to Gq, and activates the same intracellular pathway as AVPR1a and AVPR1b. It is expressed in the uterus, the mammary gland, the ovary, lactotroph cells and, in the brain, in the central and basolateral amygdala, medial preoptic area, anterior and ventromedial hypothalamus, olfactory nucleus, vertical limb of the diagonal band of Broca, ventrolateral septum, anterior cingulate, and hypoglossal and solitary nuclei (Loup et al., 1991; Boccia et al., 2013). No expression has been reported in the hippocampus (including CA2 and CA3), parietal cortex, raphe nucleus, nucleus ambiguus or pons (Boccia et al., 2013).

Interestingly, there is some evidence for the formation of heterodimers between AVPR1a, AVPR2, and OXTR (Cottet et al., 2010). Formation of such heteromeric receptors could have important in vivo implications and possibly confound the interpretation of hormonal and neuromodulatory actions (Cottet et al., 2010; Devost and Zingg, 2003). Despite the highly conserved protein structure of the peptides and of their receptors, it is noteworthy (particularly when extrapolating to humans findings from research done on animals) that their corresponding gene structure and spatial pattern of expression presents species-specific variations, as shown, for instance, in the sequence and brain expression of AVPR1a in monogamous and non-monogamous vole species (Hammock and Young, 2004; 2005). The variation on the gene structure of OXTR and AVPR is mainly in the form of genetic polymorphisms, which have been putatively associated, inter (Young et al., 1999) and intraspecifically (Knafo et al., 2008; Young et al., 1999), with promotor activity (Tansey et al., 2011) and brain expression (Knafo et al., 2008; Young et al., 1999) of the receptors in the brain.

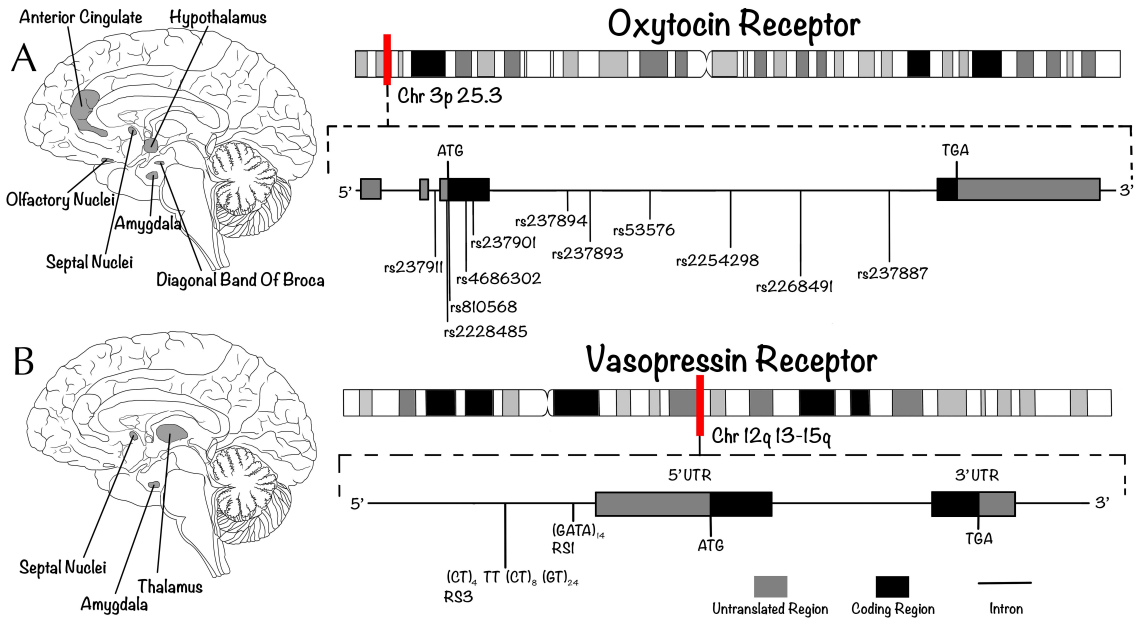


Figure 6: OXTR and AVPR are G-protein coupled receptor expressed in key structures of the brain. Their genes present characteristic polymorphisms associated with differences in human social (and pathological) behaviors. **(A)** In the human brain, OXTR is expressed in the basolateral amygdala, the anterior and ventromedial hypothalamus, the olfactory nucleus, the diagonal band of Broca, the septal nuclei, and the anterior cingulate (left). Its gene (right) is located on chromosome 3p25.3 (approximate position indicated by red vertical line). It contains four exons and three introns, which include several known SNPs. **(B)** AVPR1A is expressed in the septal nuclei, the thalamus and the basal amygdaloid nucleus (left); the gene encoding this receptor (right) is located on chromosome 12q14 (approximate position indicated by red vertical line). As in the case for OXTR, it contains an intron before the exon that encodes the seventh transmembrane domain protein. The schematic includes the characteristics SNPs (in the case of OXTR) and SSRs (in the case of AVPR) that are reviewed in this article.

3.2.2 Types of polymorphisms

A genetic polymorphism is the occurrence of two or more genetic variants for a particular locus in the DNA sequence (i.e., alleles) within a population. Polymorphisms can be classified depending on the types of changes observed. One of the most common is a change at the level of single nucleotides (single nucleotide polymorphism,

SNP). SNP's located in coding regions of the DNA can in turn be classified as synonymous and nonsynonymous, depending on whether or not they alter the amino-acid sequence of the resulting protein, respectively. Finally, a non-synonymous SNP can change the protein's amino-acid sequence (i.e., missense mutation) or result in a truncated protein (i.e., nonsense mutation). SNPs located in noncoding regions can affect the levels and spatial expression of a gene, and can therefore also alter gene action (Shastry, 2002). Several SNPs can occur simultaneously at different locations within a given gene and will therefore tend to segregate together. These various SNPs are called haplotype blocks, or simply haplotypes. The occurrence of these blocks is useful for investigating the genetic bases of common diseases, as it can reduce the number of variants from 500,000 to nearly 350,000 (Plomin, 2008).

Another common type of polymorphism is microsatellites or tandem simple sequences repeats (SSRs). In this case, alleles differ in the number of SSR repeats present. Despite the fact that most of the microsatellites in the human genome are located in non-coding regions, and are thus usually considered to be evolutionary neutral, they can nevertheless cause phenotypic differences (Li et al., 2004), including for neurological disorders in humans (Cummings and Zoghbi, 2000).

3.3 Inheritance studies on human behavior

Studies on behavioral genetics have consistently shown that heritability (the proportion of phenotypic variance attributable to genotypic variance) plays a role in determining human complex social traits. Yet, this heritability rarely exceeds 50%, which highlights an important fact: environmental and epigenetic factors almost always account for more than half of the variance (Plomin, 2008). However, behavioral genetic research has shown that many behaviors, including social behaviors and

their deficits, which are typically considered to depend exclusively on environment, show clear genetic influences (Plomin, 2008). According to Plomin (2008), behaviors can depend on genotype in (at least) three ways: (i) passively, where children inherit from their parents an environment that is correlated with the children's genetic propensities; (ii) evocatively, where children evoke reactions from other people on the basis of their genetic propensities, and (iii) actively, by selecting, modifying and constructing experiences correlated with their genetic propensities. Thus, a child inheriting a depressive "allele" could also inherit a depressor environment from her depressive parents (passive); act apathetically, making others avoid her (evocative); and develop cognitive biases, viewing the life through grey glasses and creating, in this way, her own depressing reality (active). These genetic influences on environmental measures have been called genotype-environment correlations (Plomin, 2008). In addition, there exist genotype-environment interactions, where the effect of environment on some traits depends on the genotype and, conversely, the effect of a given genotype depends on the environment. Phenotypes are, in this way, more than the sum of independent genetic and environmental factors. Undoubtedly, research has shown that variations in socially important behaviors are sensitive to variations in genotypes (Plomin, 2008).

3.4 Linkage and association studies

In classical linkage studies, heritability is measured on the basis of correlations found among relatives. These studies typically analyze large pedigrees of families presenting a characteristic trait, and, ideally, include monozygotic (MZ) and dizygotic (DZ) twins. Because MZ twins have exactly the same genotype, whereas DZ twins on average share only about 50% of their genetic material, if genetic differences account

for phenotypic differences, then MZ twins should exhibit higher correlation than DZ twins in any particular trait (Cesarini et al., 2008, Plomin, 2008). Furthermore, if MZ and DZ twins live in a shared environment, then these correlations are useful for estimating the relative influence of genetic vs. environmental factors on phenotypic variation (Cesarini et al., 2008; Plomin, 2008). A classical success story of this approach is in the analysis of a five-generation pedigree of a family with members afflicted by Huntington disease, and which eventually lead to the identification of the CAG repeat associated with most cases of the disease (MacMillan et al., 1993; Plomin, 2008). Alternatively, sib-pair linkage analyses can be carried out, in which case a small number of relatives are considered, but in a large number of families. In both approaches, the researcher is interested in looking for a marker that is co-inherited with the trait of interest. However, these approaches cannot detect associations in which genes have small effects, as is the case for most complex behaviors or disorders (Risch, 2000; Plomin, 2008). Single-gene disorders tend to be rare, whereas many important multifactorial traits or diseases, including those affecting human social behavior, autism, depression, and schizophrenia, occur at much higher frequencies in the population, and also show high heritabilities (Risch, 2000; Plomin, 2008).

Association studies, rather than relying on the incidence of a disease within families (and of an associated marker), test whether the frequencies of alleles or genotypes differ between groups such as affected individuals vs. controls, or among individuals with extreme scores on a quantitative trait (Sham et al., 2000; Plomin, 2008), testing whether specific alleles correlate with increased or decreased scores on trait or prevalence of a disease (Plomin, 2008). These studies have classically been restricted to a few "candidate" genes (see, for instance, Knafo et al., 2008; Rodrigues et al.,

2009; Luo et al., 2005; de Waal et al., 2008; Thompson et al., 2011). The weight of a given genetic factor is then quantified, for instance, using the odds ratio, which is calculated as the ratio of the frequency of an allele in affected individuals divided by the frequency of this same allele in the control sample. By this measure, an odds ratio of 1.0 means that there are no difference in allele frequencies between affected individuals and controls, whereas an odd ratio of 3.0, for instance, is considered to be a large effect. (Plomin, 2008; Brookfield, 2010).

3.5 Pitfalls of linkage and association studies

In any association study, it is critical that the groups be well-matched in terms of ethnicity, gender, and age (Plomin, 2008; Brookfield, 2010). If not, the study could identify spurious factors, which are simply due to differences between the groups (Plomin, 2008; Brookfield, 2010). For instance, if a disease is more prevalent in one of two ethnically different populations, then any genetic variant that is more common in one of these populations will show an association with a disease simply because of differences in ethnicity. This could be the case for some reports reviewed here, which show inconsistent results but were based on different ethnics groups.

3.6 Quantitative traits loci (QTL)

In a classical monogenic approach, the presence or absence of the trait follows a dichotomous distribution. In the case of qualitative diseases (diseases where the trait is either present or not), a single allele would be necessary and sufficient (Plomin, 2008; Brookfield, 2010). Currently, around one thousand single-gene, qualitative disorders are known (Plomin, 2008). However, when the penetrance is very low, it is virtually impossible to map genes using pedigrees (Plomin, 2008; Brookfield, 2010);

more importantly, most human social and pathological behavior is better described as a qualitative and continuous, rather than as a discrete, qualitative variable (Plomin, 2008).

A polygene (or quantitative gene) is a group of usually pleiotropic genes, that together influence a phenotypic trait, leading to additive, non-epistatic effects (Plomin, 2008; Brookfield, 2010). In this way, a quantitative trait is a phenotype that varies in degree and can be attributed to polygenic effects, i.e., the product of two or more genes. Each gene is inherited in a Mendelian fashion, but the additive effect of several genes results in phenotypes that approach a normal distribution within the population (Plomin, 2008; Brookfield, 2010). For instance, if n is the number of loci involved, then the coefficients of the binomial expansion of (a + b)2n will give the frequency of distribution of all n allelic combinations (Thurnpenny, 2004; Plomin, 2008). For a sufficiently high n, this binomial distribution will begin to resemble a normal distribution (Plomin, 2008). When environmental effects are included in the model, we enter into the world of multifactorial traits or disorders, where the presence or absence of the disease is influenced by many genetic differences and, of course, by the environment (Plomin, 2008). This quantitative traits loci approach has strong implications on how we conceive human social and pathological behavior, because there must be a continuum of genetic risk, in which diagnosed subjects could be extreme cases that differ quantitatively, but not qualitatively, from normal cases (Plomin, 2008).

3.7 Role of OXTR and AVPR in social behaviors

3.7.1 Studies in rodents

Reports implicating OXT and AVP in the control of mammalian social behavior first appeared in the early 90's. For instance, Insel and Shapiro (1992) showed that the distribution of OXTR in the brain of voles differed between monogamous (prairie) and polygamous (montane) species. Likewise, AVPR distribution also differed between these species, most likely because of differences in promoter structure (SSRs polymorphisms; Young et al., 1999). Most interestingly, intracerebroventricular injections of AVP increased affiliative behaviors in the monogamous prairie vole, but not in the promiscuous montane species (Young et al., 1999), and transgenic mice expressing the prairie vole variant of AVPR also showed increased affiliative behaviors in response to AVP injections (Young et al., 1999). Collectively, these findings suggest that the different pattern of AVPR expression is responsible for the different behavioral responses to AVP in these two related species of voles.

Since these early studies in animal models, a large amount of research has been carried out on these two so-called "prosocial neuropeptides", most of them using the powerful techniques available in rodent models (Nakajima et al., 2014; Yan et al., 2015; Sala et al., 2015). Regarding OXT, for instance, Nakajima et al. (2014) demonstrated that blocking neurotransmission from medial prefrontal cortex (mPFC) interneurons that express OXTR (specifically, somatostatin neurons) caused female mice to lose interest in male mice during the sexually receptive phase of the estrous cycle (Nakajima et al., 2014). Similarly, behavioral despair (a behavior associated with depression) was shown to enhance OXT synthesis and secretion in the paraventricular nucleus, supra-optic nucleus, frontal cortex, amygdala, and hippocampus, as

well release from the posterior pituitary into the blood (Yan et al., 2015). Finally, and as way of example, OXTR null mice (OXTR -/-) display autistic-like deficits in social behavior and increased aggression, in addition to a resistance to changes in a learned pattern of behavior (comparable to restricted interests and repetitive behavior in autism; Sala et al., 2015). Furthermore, this study suggests that intracerebral administration of both OXT and AVP reverts the social and learning defects by acting on AVPR1a receptors (Sala et al., 2015). Insights have also been obtained into the possible functionality of some of the AVPR alleles, including the impact of SSRs on the pattern of AVPR expression in the brain (Hammock et al., 2005; Young et al., 1999), transcriptional regulation (Hammock and Young, 2005; 2004) and promotor activity (Tansey et al., 2011), all of which are associated with differences in social behaviors like affiliation (Insel, 2010;), aggression (Beery et al., 2008), mating (Young et al., 1999; Caldwell et al., 2008) and anxiety-like behaviors (Neumann and Landgraf, 2012; Hammock et al., 2005).

Regarding psychiatric models, OXT administration, for instance, has been shown to mimic the effects of antipsychotic drugs. For example, subcutaneous OXT administration in rats reduces prepulse inhibition, a behavioral measure of sensorimotor gating in which a startle response to a stimulus is reduced when it is preceded by a weaker stimulus (Feifel and Reza 1999; Feifel et al., 2012). These results are relevant since prepulse inhibition is a neurological phenomenon commonly seen in schizophrenic patients (Ratajczak et al., 20013; Powell et al., 2009) suggesting, thus, that OXT could produce antipsychotic-like central effects.

3.7.2 OXTR and AVPR genetics in human social and pathological behavior

These results suggest that OXT and AVP are involved in regulating a wide range of social behaviors. However, results from rodents are not easy to extrapolate to primates, including humans. For instance, Fink et al. (2006) found that, in primates, the polymorphisms upstream of the AVPR1a, which have been associated with differences in social bonding, are evolutionarily distinct from those present in rodents. In addition, studies on the role of variations in the genes encoding OXT and AVP on social human behavior have largely been uninformative (Meyer-Lindenberg et al., 2011). Nevertheless, a substantial body of evidence obtained through linkage and association studies now implicates sequence polymorphisms in their receptors in the modulation of a growing list of human behaviors (Meyer-Lindenberg et al., 2011).

3.7.3 Monogenic approach

Even though the genes coding for AVPR and OXTR share a common ancestor (Caldwell et al., 2008; MacDonald and MacDonald, 2010; Insel, 2010), they are located on different chromosomes. In humans, the 17Kb OXTR is located on chromosome 3p25.3 and contains four exons and three introns, one of them containing more than twelve known SNPs (Inoue et al., 1994). In the case of AVP, the genes encoding the 3 receptors (AVPR1a, AVPR1b, and AVPR2) are located on chromosomes 12q14, 1q32 and Xq28, respectively. As is the case for OXTR, all contain an intron before the exon that encodes the seventh transmembrane domain (Thibonnier et al., 1996; Sugimoto et al., 1994; Seibold et al., 1992). OXTR and the three AVPR genes present characteristic polymorphisms that have recently been associated with variations in social behaviors, suggesting that they could be part of the genetic sources

for the heterogeneity observed in social traits and psychiatric disorders (see below).

TABLE 1 | OXTR and AVPR polymorphisms associated with social and pathological behaviors.

Trait	Gene	Polymorphism studied	Species/Ethnicity	Associated Genotype or Haplotype	Behavioral phenotype	Brain/Biological phenotype	Interactions with other gene or variable	References
Aggression	AVPR1b	rs35369693	147 Europeans, 10 African Canadians, 2 East Asians, and 18 of mixed ethnicity	C allele	Increased aggression	–	Haplotypes with rs28676508	Zai et al., 2012; Luppino et al., 2014
Autism	AVPR1a	RS1	Irish	Short allele	Autism	Increased activity in the left amygdala in a non-clinical sample	Possibly reduced transcription of AVPR1A	Tansey et al., 2011
Autism	AVPR1a	RS1; RS3	94 Caucasian, 7 African-American, 8 Asian-American and 6 Hispanic (17 females)	RS3	In linkage disequilibrium in autism (not significant after bonferroni correction)	–	–	Kim et al., 2002
Autism	AVPR1a	RS1; RS3 and AVR	Not specified ($n = 128$ probands, 3 females)	AVR	Transmission desequilibrium in autism; Moderate linkage desequilibirum and association with some scores related to autism	–	Haplotypes involving alleles of the three polymorphisms	Yirmiya et al., 2006
Autism	OXTR	rs2254298 and rs53576	57 Caucasian trios	rs2254298 G	Overtransmited in autism	Phenotypic heterogeneity	Haplotypes with an as yet unidentified susceptibility variant	Jacob et al., 2007
Autism	OXTR	rs2254298, rs53576, rs237893, rs237894, rs237911, rs237901, rs810568, rs2228485	195 Chinese Han family trios (21 females)	rs2254298 A and rs53576 A; haplotypes involving rs53576	Displayed a preferential transmission in Autism	–	–	Wu et al., 2005
Autism	AVPR1b	rs2254298, rs53576, rs237893, rs237894, rs237911, rs237901, rs810568, rs2228486	89 Swedish, 89 Belgians in the patients sample (122 females); 88 Swedish and 89 Belgiansin the control sample (117 females)	G allele for s3 in Swedish sample; G allele for s5 in the Belgian sample; A-T-C-A-G Haplotype in both samples	Recurrent major depression; Overrepresented in controls; Overrepresented in controls	–	Ethnicity	Van West et al., 2004
Depression	OXTR	rs2254298	92 Caucasian adolescent girls	Heterozygous	Highest levels of symptoms of depression, physical and social anxiety, if they presented history of maternal depression	–	Maternal history of recurrent major depressive disorder	Thompson et al., 2011
Depression	OXTR	rs2254298	Not specified (first cohort composed of 1983 pregnant females)	G Homozygous	Overrepresented in depressed mothers and families (including fathers)	Lower salivary OXT	–	Apter-Levy et al., 2013
Depression	OXTR	rs53576 (G/A)	413 Caucasian, 18 Asian, 5 Maori/Pacific Islander, 2 Aboriginal and 1 other (261 females)	A Carriers	Higher depressive symptoms if mother has been depressed	–	Association between maternal depression in early childhood and youth depressive symptoms in adolescence	Thompson et al., 2014

(Continued)

Figure 7: Neuropeptides polymorphism associated with social and pathological traits.

TABLE 1 | Continued

Trait	Gene	Polymorphism studied	Species/Ethnicity	Associated Genotype or Haplotype	Behavioral phenotype	Brain/Biological phenotype	Interactions with other gene or variable	References
Depression	OXTR	rs53576 (G/A)	167 White, 34 Black, 23 Asian, 17 Arab, 16 South Asian, 6 South East Asian, 4 Latin American, 17 other (213 females)	G Carriers	Increased depressive symptomatology if suffered childhood maltreatment	–	Childhood maltreatment	McQuaid et al., 2013
Depression	OXTR	rs53576 (G/A)	128 White female	G Carriers	More emotionally sensitive (lower self-esteem) in response to social ostracism	Altered blood pressure and cortisol levels following rejection for G homozygous	–	McQuaid et al., 2015
Depression	OXTR	rs53576; rs2254298	99% Caucasians (71% females; $n = 268$)	No association found	Antidepressant treatment resistance, response or remission	–	Cyclooxygenase-2 (COX-2), was associated with antidepressant treatment resistance	Mendlewicz et al., 2012
Depression	OXTR	rs53576	43 clinically depressed and 42 healthy female subjects	Exon 1 and 2 methylation patterns	Decreased methylation state were associated with depression	Regulate DNA methylation throughout the whole genome, depending on early environment	Specific genotypes can regulate DNA methylation	Chen et al., 2015
Empathy	OXTR	rs2268491; rs2254298	Non-clinical Chinese subjects (46 males, 55 females)	CT higher than CC; CT higher than TT	Cognitive empathy	–	Gender- dependent	Wu et al., 2012
Empathy	OXTR	rs237887; rs4686302	Non-clinical Chinese subjects (46 males and 55 females).	A Carriers; T carriers	Higher Emotional empathy	–	Gender-dependent	Wu et al., 2012
Empathy	OXTR	rs53576 (G/A)	35% Caucasian, 41% Asian, and 24% others. (n =192)	G Homozygous	Lower levels of dispositional stress reactivity, exhibit greater empathy	Lower heart rate reactivity to a startle anticipation task	–	Rodrigues et al., 2009
Empathy	OXTR	rs53576 (G/A)	No specified ($n = 1532$)	G Homozygous; A Homozygous	Greater in-group bias in implicit attitudes; Less motivation to reduce out-group member pain	Higher ACC and SMA activity; Higher NAc activity	–	Luo et al., 2015a
Empathy	OXTR	rs53576 (G/A)	Chinese students ($n = 1536$, 710 females)	G Homozygous	Showed a stronger association between empathy and interdependence	Insula, amygdala and superior temporal gyrus	Associated with racial in-group	Luo et al., 2015b
Harm avoidance	OXTR	rs53576 (G/A)	Chinese Han ($n = 290$, 154 females)	A Homozygous	Increased harm avoidance	Smaller amygdala volumes bilaterally Reduced resting-state functional coupling between the prefrontal cortex and amygdala bilaterally in females	Gender- dependent	Wang et al., 2014

(Continued)

TABLE 1 | Continued

Trait	Gene	Polymorphism studied	Species/Ethnicity	Associated Genotype or Haplotype	Behavioral phenotype	Brain/Biological phenotype	Interactions with other gene or variable	References
Neuroticism	OXTR	rs53576 (G/A)	Chinese Han ($n = 82$, 45 females)	G Carriers; AA	MPI neuroticism scores correlated with DAT and OXT interaction (AA)	Lower striatal DAT availability; Negative correlation between DAT availability and OXT levels in G carriers	–	Chang et al., 2014
Prosocial Behavior	AVPR1a	RS1 and RS3	Non specified ($n = 203$, 101 females) ($n = 15$ for post mortem studies)	Long (327–343 bp) alleles	Larger allocations in a dictator game and higher scores in self-reported altruism scales	Increased mRNA hippocampal levels	–	Knafo et al., 2008
Prosocial Behavior	OXTR	rs53576 (G/A)	108 European - Americans males	G Homozygous	Higher trust behavior	–	–	Krueger et al., 2012
Schizophrenia	AVPR1a	RS3	Russians (291 patients, 49% females; 302 control group,59% females)	327 bp allele	Schizophrenia's negative symptoms; facial affect recognition	–	–	Golimbet et al., 2015
Schizophrenia (Attention)	AVPR1a	RS1; RS3	Non specified ($n = 113$, 75 female)	Long alleles (>325 bp)	Promoter repeat region in partially molding social behavior in both animals and humans	Greater levels of prepulse inhibition	Gender (stronger association in male)	Levin et al., 2015
Stress	OXTR	rs53576 (G/A)	White (74.9%), Black (8.92%), Hispanic (8.92%), and other (52.3% female)	A carriers; G Homozygous	Increased Post traumatic symptoms in negative social environment, independently of economic stress; Elevated post traumatic symptoms in negative social environment if economic stress exists	–	Economic stress	Thompson et al., 2011
Stress	OXTR	rs53576 (G/A)	173 European participants, 15 mixed, and 6 "other." All males	G carriers	Attenuated anticipatory stress response if they received social support	Lower cortisol responses under stress after social support	Social support	Chen et al., 2011
Stress	OXTR; AVPR1a	rs53576; RS1	37% Asian-American, 2% African-American, 23% European-American, 16% Latin American, 7% mixed ethnicity and 15% other. ($n = 172$, 60% females)	G Homozygous; 320 bp	Women with higher poststressor OXT levels reported more positive affect feelings; Men with high levels of poststressor AVP reported more anger and hostility feelings	–	Gender and poststressor levels of OXT and AVP	Moons et al., 2014

The majority of polymorphisms studied in OXTR gene are SNPs, and some of them appear to be strongly associated with variability in social traits. One of the most studied is rs53576, which consists of a G to A change within the third intron of OXTR. In the case of AVPR1, the most studied polymorphisms are SSRs, such as RS3 and RS1, the former located upstream of the transcription start site and the second located in a coding region 553 bp downstream from the start site (Meyer-Lindenberg et al., 2011). Despite such associations, their impact on gene expression and function are still unknown. In the case of AVPR1b, SNP polymorphisms have recently been discovered, which have been associated with social and psychiatric traits such as protection to major depression (West et al., 2004) and childhood-onset mood disorders (Dempster et al., 2007; Zai et al., 2012). To date, AVPR2 gene polymorphisms have not been associated with any social trait; yet their relation with diabetes is well established (Spanakis et al., 2008).

Prosocial behavior. An interesting line of research has shown that humans display several (pro)social behaviors, including trust, generosity and altruism. These result are of great interest to fields such as Neuroeconomics (Fehr et al., 2007; Ebstein et al., 2009), because the assumption of neoclassical economics is that individuals always act to maximize a utility function which depends exclusively on their own material gain (Fehr and Schmidt, 1999). Because of a growing body of evidence that does not support this hypothesis, this utility function has had to be modified to include other-regarding preferences (Fehr and Schmidt, 1999; Dufwenberg and Kirchsteiger, 2004; Rabin, 1993)

Two independent studies of monozygotic and dizygotic twins, one in Sweden and one in the United States, have shown that there is a significant heritability component to prosocial behaviors. Thus, Cesarini et al. (2008) estimated that correlations in

trust behavior, as measured in a classical Trust game, are 0.13 and 0.25 in the case of North Americans and Swedish MZ twins, respectively, vs. 0.07 and 0.01 in both populations of DZ twins. The same applies for trustworthiness: MZ were correlated in 0.26 (North Americans) and 0.29 (Swedish), vs. 0.06 (North Americans) and 0.18 (Swedish) for DZ. The heritability estimate is 10% (with a 4-21% confidence interval) in the U.S. study and 20% (with a 3–38% confidence interval) for the Swedish study (Cesarini et al., 2008), with unshared environmental variation having a greater impact on phenotypic variation than genetic variation. This group has also shown that strategies and fundamental economic preference parameters are moderately heritable, with estimates for heritability ranging from 18 to 42% (Cesarini et al., 2009) and, importantly, suggest that these traits have a polygenic architecture, with the heritable variation being explained by many genes with small effects (Benjamin and Cesarini, 2012).

Several association studies are also beginning to shed light on the molecular basis of variations in these so-called "prosocial" behaviors (Krueger et al., 2012; Kosfeld et al., 2005; Zak et al., 2007; Radke and Bruijn, 2012). In one of the first reports on the field, Knafo et al. (2008) classified the AVPR1a RS3 SSRs of participants into short (308-325 bp) and long (327-343 bp) versions, and found, using the Dictator game, that subjects with the short version allocated significantly fewer tokens to others than did participants with the long versions. This was confirmed and validated using the family-based association test and two self-report scales (the Bardi-Schwartz Universalism and Benevolence Value-expressive Behavior scales). Interestingly, long AVPR1a RS3 repeats were associated with higher AVPR1a human post-mortem hippocampal mRNA levels than were short RS3 repeats (Knafo et al., 2008). Along a similar line, Krueger et al. (2012) reported that individuals homozygous for the

rs53576 G allele show higher trust behavior (not just a general increase in trustworthiness or risky behaviors) than do A carriers (i.e., individuals presenting at least one A allele in this SNP; Krueger et al., 2012). Similarly, OXT infusions increase trust (Kosfeld at al., 2005), generosity (Zak et al., 2007), and decrease adherence to social norms (Radke and Bruijn,, 2012).

On the other hand, AVPR1b alleles have been associated with aggression levels in children. Luppino et al. (2014) reported a significant relationship between the minor C allele on the SNP rs35369693 of AVPR1b and teacher-rated reactive aggression (Luppino et al., 2014). Similarly, Zai et al. (2012) found that this same SNP, as well as haplotypes containing rs35369693 and rs28676508 (both in AVPR1b), are associated with higher child aggression (subjects that scored at or above the 90th percentile on the aggression subscales of both the Child Behavior Checklist and the Teacher's Report Form; Zai et al., 2012).

Empathy. Empathy is a social behavior defined as a capacity to share and understand the feelings of others (de Vignemont and Singer, 2006). Rodrigues et al. (2009) reported that subjects homozygous for the G allele of OXTR SNP rs53576 exhibit greater empathy, in comparison with A carriers, as measured by the "Reading the Mind in the Eyes" task. Luo et al. (2015) compared the behavior and the fMRI responses of individuals when they observed painful or non-painful stimulations of others subjects, who were categorized by racial ingroup/outgroup criteria. They found that individuals that were homozygous for the G variant showed increased activity in the anterior cingulate cortex (ACC) and supplementary motor area (SMA) in response to racial ingroup members' pain, as compared to individuals homozygous for the A variant. By contrast, individuals homozygous for the A allele showed increased activity in reward system structures, specifically in the nucleus

accumbens, when they watched a racial outgroup member in pain. This opposite response also predicts the participants' attitudes and behavior: the racial ingroup bias in ACC/SMA activity was positively correlated with participants' racial ingroup bias in implicit attitudes (specifically, they made more associations between Asian vs. Caucasian faces and concepts of "good" vs. "bad", as measured by the Implicit Association Test), whereas nucleus accumbens activity showed a negative correlation with participants' motivations to reduce racial outgroup members' pain.

Noticing the different results obtained between Eastern vs. Western subjects in studies on empathy, this same group also investigated whether OXTR rs53576 SNP interacts with the interdependence trait (i.e., how people view themselves in relation to others) to modulate human empathy. They found a stronger association between interdependence and empathy in G carriers as compared to A homozygotes (Luo et al., 2015). In addition, they used fMRI to measure the neural responses elicited by the suffering of others in A/A vs. G/G homozygotes, finding stronger associations between interdependence and empathic neural responses in G/G compared to A/A genotypes in brain structures such as the insula, amygdala, and the superior temporal gyrus (Luo et al., 2015).

Other studies have investigated the association between other common OXTR SNPs and empathy. Emotional empathy is defined as the ability to respond with an appropriate emotion to another's' mental states (Simone et al., 2009; de Waal, 2008). Wu et al. (2012) found an association between emotional empathy and OTXR variation at the rs237887 SNP (with A allele subjects scoring higher than those with the G allele), and the rs4686302 SNP (with T allele subjects scoring higher than those with the C allele). By contrast, cognitive empathy, the capacity to understand another's perspective or mental state (Simone et al., 2009; de Waal, 2008), showed associations

with polymorphisms SNPs rs2268491 (with carriers of the CT variant scoring higher than those of the CC) and rs2254298 (with CT scoring higher than TT; Wu et al., 2012).

Autism. Autism is a neurodevelopmental disorder characterized by abnormalities in social relationships, communication deficits, and restricted interests (Plomin, 2008), with a prevalence of 3-6 cases per 10,000 people, being four times more common in males than in females (Plomin, 2008). Although the risk of autism in offspring of autistic parent(s) is small (about 5%; Plomin, 2008), it is nonetheless 100 times greater than the rate of autism in the general population (Plomin, 2008). However no specific genes have reliably been implicated with this disorder (Freitag, 2007; Plomin, 2008) and the corresponding linkage studies have shown poor replicability (Trikalinos et al., 2006; Plomin, 2008), although there are reports of a QTL on chromosome 7 (7q31-q33) that shows some association with the disease (Plomin, 2008).

A couple of studies have investigated the association between SNPs rs53576 and rs2268498 and autism, with rather confusing results. In one of two family-based association tests (FBAT; Horvath et al., 2001), Wu et al. (2005) found, in a study involving Chinese Han individuals, a significant association between autism and individual variants of OXTR SNPs rs53576 and rs2254298, as well as with haplotypes involving rs53576 (specifically, A-A-T-A in rs53576, rs2254298, rs2228485 and rs237911, respectively). By contrast, Jacob et al. (2007), using Caucasian individuals, found a significant association between autism and polymorphism at the rs2254298 SNP, but not at rs53576. Moreover, they found that the overtransmission (i.e., a higher likelihood that the risk allele is transmitted to the affected child, as compared to the non-risk allele) of the G allele was more strongly associated with the autistic disorder group, again in contrast to the study involving Han patients (Wu

et al., 2005), which, as noted above, reported the overtransmission of the A allele (Jacob et al., 2007). In addition, there are some reports associating the short alleles of SSR RS1 in AVPR with autism: Tansey et al. (2011), for instance, found that the short alleles of RS1 show a weak association with autism in an Irish population, consistent with previous studies (Yirmiya et al., 2006; Wassink et al., 2004; Kim et al., 2002).

Schizophrenia. Schizophrenia is perhaps the most studied disorder of behavioral genetic research in psychopathology (Plomin, 2008). Its typical symptoms include delusions, hallucinations (especially auditive), disorganization of speech and behavior, and negative symptoms (i.e., loss of normal processes) like flat affect and avolition (Gottesman, 1991; Plomin, 2008). It is estimated that about 1% of the population is afflicted by the disorder, which presents high heritability (about 48% in identical twins; Gottesman, 1991; Plomin, 2008). Interestingly, although it runs in families, the particular subtype (catatonic, paranoid, disorganized) does not (Plomin, 2008).

There exist a small but significant association between schizophrenia and polymorphisms in the Neuregulin I gene (Lewis, 2003; Plomin, 2008), which is involved in the development of the nervous system (Harrison and Law, 2006; Plomin, 2008). Although there is currently no evidence of an association between OXTR and AVPR polymorphisms and schizophrenia, there are reports of associations between OXTR and AVPR polymorphisms and traits associated with the disease, such as attention (the main cognitive domain that is impaired in this disorder; Ratajczak et al., 20013; Powell et al., 2009) and negative symptomatology. Golimbert et al. (2015) showed that the AVPR1A 327bp SSR is associated with negative symptoms (as measured using the Positive and Negative Symptoms Scale, PNSS), and tended to be linked with patient facial affect recognition, probably impacting social phenotypes

of schizophrenia. In addition, Levin et al. (2015) used a robust family-based strategy to show that longer RS3 alleles were associated with greater levels of prepulse inhibition, a classical paradigm designed to measure attention. This association was more prominent in males.

Depression. Depression is a state of low mood and aversion to activity, with characteristic symptoms including anhedonia, overeating or loss of appetite, insomnia, excessive sleep, fatigue, aches, pains, digestive problems and reduced energy (McGuffin and Katz, 1985; Plomin, 2008). Subjects suffering severe depression may contemplate, attempt to, or commit suicide (McGuffin and Katz, 1985; Plomin, 2008), emphasizing its societal importance. The familial risk of major depression and for bipolar disorder is about 9%, as compared to 3% and 1% in control samples for depression and bipolar disorder, respectively (McGuffin and Katz, 1985; Plomin, 2008).

Some reports have shown heritability of protective alleles and depression-related traits. Consistent with an interpersonal perspective on depression (Joiner and Coyne, 1999), Apter-Levi et al. (2013) found that in the families of depressed mothers, salivary OXT levels was lower in mothers, children and (curiously) also in fathers, as compared to control families. Children, in addition, had lower empathy and social engagement levels, and 61% of the children displayed axis I disorders, mainly anxiety and oppositional defiant disorder, compared to 15% of children of nondepressed mothers. The OXTR rs2254298 SNP homozygous for the G variant was overrepresented in depressed mothers and their families (including fathers), and correlated with lower salivary OXT. The presence of a single A allele in this SNP in depressed mothers markedly decreased the risk of child psychopathology (Apter-Levy et al., 2013).

Another study found that youth possessing at least one A allele of the OXTR rs53576 SNP, whose mothers had had a history of depression (as measured at age 15, using the SCID-I scale), exhibited the highest levels of depressive symptoms at age 15 (as measured by the Beck Depression Inventory II; BDI-II), showing that SNP rs53576 acts as a moderator variable in the transmission of maternal depression from mothers to their children (Thompson et al., 2014). In addition, under high levels of childhood maltreatment (as reported by the Childhood Maltreatment Questionnaire) only carriers of the G allele of SNP rs53576 presented increased depressive symptomatology (as measured by BDI) when compared to those with the AA genotype (McQuaid et al., 2013). Moreover, carriers of the G allele were more emotionally sensitive (lower self-esteem) in response to social ostracism (as measured by the Social Ostracism and Mood Scale), and showed altered blood pressure and cortisol levels following social rejection induced by the Cyberball task, a well-established computerized game used to induce feelings of social rejection (Williams et al., 2000; McQuaid et al., 2015).

There are also reports on the putative role of polymorphisms in AVPR genes and mood disorder spectrum. In a Swedish and a Belgian study, for instance, the haplotype defined by alleles A-T-C-A-G for the AVPR1b SNPs s1-s2-s3-s4-s5 was significantly over-represented in controls vs. patients with depression (van West et al., 2004). Despite these associations, there are reports that, at least for OXTR, polymorphisms do not alter antidepressant treatment resistance, response or remission, nor are they associated with variations in the inflammatory pathways that have been reported to play a role in antidepressant efficacy, such as cyclooxygenase-2 and OXTR SNPs rs53576 and rs2254298 (Mendlewicz et al., 2012).

Stress. Thompson et al. (2011) found, in a sample of Caucasian girls, that subjects who were heterozygous for the OXTR rs2254298 polymorphism (presenting A-G

substitutions) and had high early adversity showed the highest levels of depression, and physical and social anxiety. Another study showed influences of the rs53576 SNP on stress: Thompson et al. (2013), for instance, found that individuals homozygous for the rs53576 G allele that have been exposed to a negative social environment (specifically have been suffering high economic stress, measured using a Likert scale created for the study) showed elevated post-traumatic stress symptoms after the 9/11 events, as measured using the post-traumatic stress disorder (PTSD) Checklist. The same risky G allele was associated with lower cortisol responses to stress after social support, compared to individuals with the same genotype, but receiving no social support (Chen et al., 2011). In addition, this G allele and the AVPR 320 bp RS1 SSR seem to have a gender specific interaction with cortisol plasmatic levels: thus women, but not men, with high levels of poststressor OXT and the GG genotype, felt the most positive affect after the stressor; by contrast only men with high levels of poststressor AVP and the 320 allele of the RS1 polymorphism reported more poststressor anger than noncarriers (Moons et al., 2014).

Harm avoidance. In females, the AA genotype in rs53576 is associated with an increase in harm avoidance and significantly smaller amygdala volumes bilaterally, especially its centromedial subregion. In addition, the AA allele is associated with reduced resting-state functional bilateral coupling between the prefrontal cortex and amygdala: in the left hemisphere this coupling was positively correlated with harm avoidance scores in female subjects (Wang et al., 2014)

Neuroticism. Carriers of the G variant of rs53576 showed lower striatal DAT availability and a negative correlation between DAT availability and OXT levels (Chang et al., 2014). Furthermore, the OXT x DAT interaction was significantly correlated with the MPI neuroticism score in the AA group (Chang et al., 2014).

3.7.4 Polygenic approach

Do OXTR and AVPR signaling pathways interact with other transmitter systems? (Sauer et al., 2013). This question has recently been approached in studies involving a couple of well known OXTR and AVPR SNPS and SSRs and polymorphisms in genes from pathways known to interact with these neuropeptide systems. A difficulty of this approach is that, mainly because of pleiotropy, we often do not have a strong hypothesis regarding which are relevant candidate pathways and genes to consider (Plomin, 2008). For such cases, an increasingly popular alternative are the genome-wide association studies (GWAS), which use a dense map of markers to genotype the entire genome. This requires scanning about 500,000 SNPs or, if selecting wisely based on haplotype blocks, about 350,000 SNPs (Cardon and Abecasis, 2003; Plomin, 2008); a valuable complement is the use of microarrays (Plomin, 2008). However, given the limited accessibility to GWAS due to their cost, and the lack of microarray chips containing the OXTR and AVPR markers reviewed here (for instance, the OXTR SNP rs53576 is not present on many currently available chips for ASD; Meyer-Lindenberg et al., 2011), selecting a couple of candidate genes and studying their interactions via polygenic association studies remains a valuable approach. In any event, there are only a few reports that have used a candidate gene polygenic approach or GWAS to investigate the effects of polymorphisms in OXTR and AVPR on social and pathological behaviors (and their functionality). These few (selected) studies are discussed below.

Autism. Nyffelet et al. (2014) found in a Caucasian population that a significant part of the risk for high functioning autism (HFA) is explained by the combination of four polymorphisms: HTTLPR (a polymorphic repeat inside the gene coding for the serotonin transporter), SNP rs6311 in the HTR2A gene, that encodes the serotonin

receptor 5-HT2A, and the rs2254298 and rs53576 SNPs in the OXTR. These data provide evidence supporting a polygenic inheritance of autism spectrum disorders (ASD), involving both the OXT and the 5-HT pathways (Nyffeler et al., 2014).

Kelemenova et al. (2010), using a sample of autistic boys in Slovakia, focused their research on several candidate gene polymorphisms for autism, specifically OXT (rs2740204), OXTR (rs2228485), GABA receptor gamma 3 (rs28431127), neuroligin (rs5916338) and reelin. The authors found only one significant association, that between autism and a higher number of GGC repeats in the (GGC)n STR polymorphism of the reelin gene, in addition to finding lower reelin levels in the blood and the brain of autistic patients.

Emotional withdrawal. Haram et al. (2015) performed an association analysis between polymorphisms at four OXT pathway genes (OXT, OXTR, AVP, and CD38) and four areas of social psychopathology (as measured by Positive and Negative Syndrome Scale) and did not find an association between rs53576 and a diagnosis of psychotic disorder, but did find an association between the A allele in this SNP and the Emotional Withdrawal traits.

Oppositional-defiant disorder. In a GWAS performed in a clinical sample of children and adolescents with Attention-deficit/hyperactivity disorder, Aebi et al. (2015) failed to find an association between oppositional-defiant disorder and the polymorphisms DRD4 exon3 VNTR (located in exon 3 of the D4 dopamine receptor gene), HTTLPR, and seven OXTR SNPs. They also performed a GWAS including the oppositional-defiant disorder dimensions and subtypes reported. Controlling for factors such as age, sex, and parental abilities, the authors did not find an association between any of the variables (Aebi et al., 2015). However, the authors did find, using

bioinformatics and literature analyses, that the proteins encoded by the 28 highest ranked genes (from the 53 included in the analysis) clustered around the beta-catenin signaling pathway, which is involved in the regulation of neurite outgrowth.

Female infidelity. A report that used GWAS identified three suggestive, but not significant, linkage areas on chromosomes 3, 7 and 20 associated with female infidelity and number of sexual partners. However, the authors did not find an association between infidelity or number of sexual partners and AVPR, a locus classically implicated in sexual behavior in mammals (Cherkas et al., 2004).

3.7.5 What do we know about the functional consequences of these polymorphisms?

Amygdala volume. For OXTR rs2254298, participants homozygous for the G allele were found to have smaller volumes of both left and right amygdala, posterior brain stem and dorsomedial anterior cingulate cortex than carriers of the A allele (Furman et al., 2011): in this same SNP, the A allele was positively correlated with bilateral amygdala volume: the larger the number of rs2254298 A alleles an individual had, the larger their amygdala volume. Furthermore, two three-single nucleotide polymorphism haplotypes, including the rs2254298 G allele, showed significant associations with a smaller bilateral amygdala volume (Inoue et al., 2010). These studies could provide a hypothesis that explains how OXTR gene variants may increase the risk of psychopathologies (Furman et al., 2011). In addition, as stated above, the AA genotype in rs53576 present significantly smaller bilateral amygdala volumes and reduced resting-state bilateral coupling between amygdala and the prefrontal cortex, specifically in women (Wang et al., 2014).

Amygdala activation. Sauer et al. (2013), studying the interaction between

OXT plasma levels and the dopamine system, specifically the common catechol-O-methyltransferase (COMT) val158met polymorphism (a polymorphism known to influence COMT activity and, consequently, DA degradation at synapses), analyzed amygdala activation following the presentation of social stimuli following placebo or OXT infusions. Their results showed no (gene) main effect and no (gene x substance) interaction, but a significant gene x gene x substance interaction. Indeed, using various social stimuli paradigms, such as the Hariri face matching task (Hariri et al., 2002a, 2002b, 2003), pictures of emotional faces from the Pictures of Facial Affect series (Ekman and Friesen, 1976), and socially relevant scenes from the International Affective Picture System (Lang et al., 2009), the authors found that, when given placebo, the effect of CD38 on bilateral amygdala activation to the presentation of social stimuli was modulated by the COMT genotype; by contrast, no such COMT genotype dependence was observed following the administration of OXT (Sauer et al., 2013). These result are consistent with the report of Baumgartner et al. (2008), who found that intranasal infusions of OXT, in addition to making subjects insensitive to unreciprocated trust (as measured with a Trust game) show reduced activation of the amygdala and the dorsal striatum when faced with social betrayal, as compared to a placebo group (see also Hurlemann et al., 2010).

Regarding AVPR, Meyer-Lindenberg et al. (2009) showed that RS1 334 bp and RS3 320 bp risky alleles (reported to have significant transmission to autistic probands) are associated with opposing effects on amygdala activation in an emotional face-matching task, a simple perceptual task previously described to robustly engage the amygdala, in which subjects must decide which one of two faces is (emotionally) identical to a target face (Meyer-Lindenberg et al., 2009). Specifically, the 334 bp RS1 allele was associated with stronger bilateral responses of the amygdala, whereas the

320 bp RS3 carriers showed a smaller activation of the left amygdala. Interestingly, there was no association between these alleles and behavioral performance (Meyer-Lindenberg et al., 2009).

Promoter activity. Both RS1 and RS3 showed differences in relative promoter activity, as measured in the human neuroblastoma cell line SH-SY5Y, with shorter repeat alleles of RS1 and RS3 showing decreased relative promoter activity (Tansey et al., 2011).

Epigenetic effects. Recent findings by Reiner et al. (2015) suggest epigenetic effects of OXTR rs53576 genotype over OXTR exon 2 methylation patterns, in females, with depressed patients presented lower levels of methylation of exon 1 moderated by OXTR rs53576 genotype (Reiner et al., 2015). This kind of findings becomes more important now that strong evidence suggest that specific genotypes can regulated DNA methylation throughout the whole genome depending on early envirnoment (Chen et al., 2015).

3.7.6 Effects of intranasal administration of OXT and AVP on psychiatric symptoms

The evidence on the effects of intranasal applications of OXT and AVP in clinical population is relatively sparse, and echoes the mixed findings for healthy population. However, the use of intranasal OXT for the treatment of psychiatric disorders shows some promise, particularly for treating symptoms involving deficits in social functioning such as autism, schizophrenia, and borderline personality and social anxiety disorders (Anagnostou et al., 2012; Bakermans-Kranenburg and van 2013; Bethlehem et al., 2013; Cardoso et al., 2014; de Berardis et al., 2013; Guastella et al., 2010a; MacDonald and MacDonald 2010; Tachibana et al., 2013a; Veening and

Olivier, 2013).

For instance, intranasal OXT has been shown to increase eye contact in individuals with ASD, possibly by increasing the saliency of social stimuli (Auyeung et al., 2015; Domes et al., 2013). Similarly, intranasal OXT in ASD patients improves the ability to recognize the social emotions of others, as measured both at the behavioral and neural levels (Aoki el al., 2014). Similarly, OXT may selectively affect the salience and hedonic assessments of socially meaningful stimuli in subjects with ASD, and thus help social attunement (Gordon et al., 2013).

In individuals with Borderline Personality Disorder (BPD), OXT administration has shown that the effects may differ depending on baseline conditions, such as the participant's representations and expectations and/or an OXT system that is not working properly in this disease. For instance, Bartz and colleagues (2011) tested whether OXT administration improves trust and cooperative behaviors in individuals with BPD vs. healthy controls using the Assurance Game (a variation of the Prisoner's dilemma; Kreps et al., 1982; Brosnan et al., 2011). They found that participants with BPD expected their partners to be less cooperative after administration of OXT vs. placebo, showing the opposite effect compared to healthy controls, where OXT infusions increase trust (Bartz et al., 2011; Kosfeld et al., 2005). In addition, intranasal administration of OXT in schizophrenic populations has shown an anxiolytic (Bell et al., 2006), antidepressant (Ozsoy et al., 2009) and antipsychotic effect (Caldwell et al., 2009; de Berardis et al., 2013; Kuehn, 2011; MacDonald and Feifel, 2012).

Surprisingly, there is very limited evidence linking OXT administration to social anxiety disorder (SAD). Guastella et al. (2009) examined the effect of OXT ad-

ministration as an adjunct to therapy for SAD, finding that patients treated with OXT showed significant differences in their ratings of speech performance and speech appearance compared to patients treated with placebo. Similarly, Labuschagne and colleagues (2010) showed, using an emotional face matching paradigm as measure, that intranasal OXT reduced amygdala reactivity to fearful faces in participants diagnosed with SAD (Labuschagne et al., 2010). In addition, the same group showed that intranasal OXT administered to SAD participants reduced cortical hyperactivity in the medial prefrontal cortex to sad faces to a level comparable to that of controls. Taken together, these studies indicate that OXT administration modulates the fear-related neural circuitry, consistent with previous research showing pro-social effects of OXT administration in healthy individuals.

Nevertheless, a recent meta-analysis examining the effect of intranasal OXT across various clinical samples (i.e., ASD, social anxiety, depression, obsessive-compulsive disorder, schizophrenia, borderline personality disorder, and post-traumatic stress disorder) found only a small to moderate effect on psychiatric symptomatology and social competence indicators (Bakermans-Kranenburg et al., 2013). Moreover, MacDonald (2013) suggested that the context-dependent and divergent effects of OXT indicate that the effects of its administration may depend on individual differences (i.e., modulation by sex, hormonal and psychiatric status, and attachment style). Nevertheless, the available evidence reveals that chronic administration of OXT causes mixed results, even within groups afflicted by the same disorder (Dadds et al., 2014; Tachibana et al., 2013b). Thus, although promising, the specific effects of OXT on psychiatric symptomatology remain unclear. In addition, the affects of acute vs. chronic OXT administration have not been thoroughly studied; thus, their relative merit for the treatment of psychiatric disorders is unknown.

Regarding AVP there is surprisingly little research on the impact of intransal AVP administration on psychiatric symptoms. Nevertheless, some studies have suggested a link between AVP levels and psychiatric disorders. For instance, cerebrospinal fluid AVP levels have been positively correlated with a life history of interpersonal aggression in individuals with personality disorders (Coccaro et al., 1998). Similarly, Pitman and colleagues (1993) investigated the effects of OXT and AVP on the emotional response of Vietnam veterans with PTSD, during personal combat imagery exercises. Responses, measured by determining skin conductance, heart rate, and electromyographic responses, showed that the group that received AVP presented higher electromyographic responses to combat imagery relative to the OXT and placebo groups, suggesting that AVP may enhance the emotional valence of events (Coccaro et al., 1998).

In healthy population, intranasal AVP has been shown to also improve social recognition. For example, Guastella and colleagues (2010) conducted a study where healthy males were asked to view faces displaying happy, neutral, or angry expressions. Using a "surprise memory test" they showed that participants who received AVP vs. placebo were more likely to remember happy and angry faces compared to neutral faces (Guastella et al., 2010b). These results suggest that AVP also enhances human recognition of emotionally-valenced faces.

The very limited research using intranasal administration of AVP (Guastella et al., 2010b; Coccaro et al., 1998) has nevertheless revealed that this neuropeptide can exert important effects on social information processing. The results discussed above largely suggest that AVP may play a role in the processing of emotionally-valenced information in both healthy and clinical population. However more research on the effects of AVP administration on clinical population is needed.

3.8 Discussion

A growing body of research has found associations between alleles of OXTR and AVPR and human social and pathological behavior, such as altruism (Knafo et al. 2008), generosity (Zak et al., 2007), aggression (Luppino et al., 2014), depression (Apter-Levy et al., 2013; Thompson et al., 2014; van West et al., 2004), empathy (Luo et al., 2015; Wu et al. 2012), autism (Kelemenova et al. 2010; Nyffelet et al. 2014; Wu et al. 2005; Jacob et al. 2007; Tansey et al., 2011), stress (Thompson et al. 2011; 2013), and schizophrenia (Golimbert et al. 2015; Levin et al. 2015). However, these results remain preliminary, and await further studies using other populations, as well as meta-analyses. As a cautionary note, a recent meta-analysis performed by Bakermans-Kranenburg and van Ijzendoorn (2013), found no significant association between any of five outcomes (i.e., biology, personality, social behavior, psychopathology and autism) and the SNPs rs53576 and rs2254298 (reviewed above), concluding that these polymorphisms failed to explain a significant part of human social behavioral diversity (Bakermans-Kranenburg and van Ijzendoorn, 2013).

Is noteworthy that both the normal and pathological behaviors covered in this review are most certainly quantitative traits involving interactions among more than one gene, and for which the phenotypes of interest are better described as a continuum. This supports the notion that the field needs to emphasize polygenic association studies where, instead of a single gene underlying a trait, there are multiple (functional) gene variants involved, each making small quantitative contributions to the final phenotype (Brookfield, 2010; Plomin, 2008). This constitutes a quantitative traits loci approach to human social behavior and psychiatric disorders.

Functionality. Another point to keep in mind is the old maxim "correlation does

not imply causation". This means that a robust result associating, for instance, a rs53576 polymorphism with a predisposition to depression does not mean that there exist a causal relationship between the two variables. However, such an association does have some inherent value: if we cannot find evidence showing a causal relationship (or if this relationship does not exist at all!), the data may still provide useful information, such as pointing to a genetic region where the relevant gene is located (Plomin, 2008). That correlation does not imply causation also raises questions about the impact of polymorphisms on gene function described in the literature. In other words: Do the OXTR's SNPs and AVPR's SSRs play a role in causing the differences in phenotypic traits? Do these variants lack functionality or do they just result from "spurious" correlations? Yet, recent research is showing that these polymorphisms are indeed associated with different volumes of some neural structures (Furman et al., 2011; Inoue et al., 2010), differential neuronal connectivity or activation (Sauer et al., 2013; Meyer-Lindenberg et al., 2009), differential expression of the receptors (Young et al., 1999; Knafo et al., 2008), plasmatic levels of the peptides (Moons et al., 2014) or promotor activity (Tansey et al., 2011). This suggests that neuropeptide receptor polymorphisms likely do have a functional role in normal and pathological human behavior. Regardless, there are also practical benefits to identifying individuals at risk even without knowing the mechanisms responsible for the deficit (Brookfield, 2010). For instance, treatments and tests are available that are offered on the basis of calculated risk (Brookfield, 2010; Plomin, 2008).

Thinking about normal and psychopathic behavior as a continuum. The notion that many genes influence complex disorders or personality traits raises the question of why such disorders are typically diagnosed as qualitative, and not as quantitative traits, given that the data support the notion that there must be a continuum of

genetic risk. This allows one to formulate two models: the liability threshold model, where risk is distributed normally and disorders occur in a qualitative fashion, when a threshold is exceeded (genetic continuum vs discrete phenotypes), vs. a model that proposes that phenotypes change continuously from normal to abnormal, with diagnosed cases being extreme cases that differ quantitatively, but not qualitatively, from normal (Plomin, 2008). The latter seems to be the best descriptor for depression and alcoholism and even for schizophrenia. It also allows finding genes that are associated with depression in afflicted patients and with mood changes in normal patients.

Finally: what do we still not know? GWAS promises to be a powerful tool for constructing the complete landscape of interacting genes but, given its current cost, will not replace research that investigates candidate genes in polygenic association studies. To date, in addition to OXTR and AVPR SNPs and SSRs, polymorphisms in COMT, CD38, HTTLPR, HTR2A and DAT are likely to be good candidates for studying the genetic bases of human social and pathological behavior. Curiously, the OXTR SNP rs53576 is not present on many currently available chips that are used for genome-wide association studies of ASD (Meyer-Lindenberg et al., 2011).

In addition, an open challenge is to determine the exact mechanism by which neuropeptides influence psychiatric symptoms. Although some reports are shedding light on the impact of different gene polymorphisms on socially relevant behaviors and their associated psychiatric disorders, the complete pathways between genes, gene interactions and behavior is still a black dark forest. As told by Chomsky, "there's a famous joke about a drunk under a lamppost looking for a pencil dropped on the ground. Somebody comes up and asks 'What are you looking for?' He says, 'I'm looking for a pencil that I dropped.' 'Well, where did you drop it?' He says,

'Oh, I dropped it across the street.' 'Well, why are looking here?' 'This is where there is light.' That's the way the sciences work. [...] If you try to move it a little further, maybe ultimately you'll get across the street."

Acknowledgments

MA and MR thanks CONICYT: Project "Anillo en Complejidad Social" SOC-1101. MA and JE thank FONDECYT (# 1141278) and Centro Interdisciplinario de Neurociencia de Valparaíso [P09-022-F], which is supported by the Millennium Scientific Initiative of the Ministerio de Economía, Fomento y Turismo.

3.9 References

Aebi M, van Donkelaar M.M., Poelmans G., Buitelaar J.K., Sonuga-Barke E.J., Faraone SV., Franke B., Steinhausen H.C., van Hulzen K.J. (2015). Gene-set and multivariate genome-wide association analysis of oppositional defiant behavior subtypes in attention- deficit/hyperactivity disorder. Am J Med Genet B Neuropsychiatr Genet. doi: 10.1002/ajmg.b.32346.

Anagnostou, E., Soorya, L., Chaplin, W., Bartz, J., Halpern, D., Wasserman, S., Wang, A.T., Pepa, L., Tanel, N., Kushki, A., and Hollander, E. (2012). Intranasal oxytocin versus placebo in the treatment of adults with autism spectrum disorders: a randomized controlled trial. Mol Autism. 3, 16. doi: 10.1186/2040-2392-3-16.

Apter-Levy, Y., Feldman, M., Vakart, A., Ebstein, R.P., and Feldman, R. (2013). Impact of maternal depression across the first 6 years of life on the child's mental health, social engagement, and empathy: The moderating role of oxytocin. The American Journal of Psychiatry, 170(10), 1161–1168.

Aoki, Y., Yahata, N., Watanabe, T., Takano, Y., Kawakubo, Y., Kuwabara, H., Iwashiro, N., Natsubori, T., Inoue, H., Suga, M., Takao, H., Sasaki, H., Gonoi, W., Kunimatsu, A., Kasai, K., and Yamasue, H. (2014). Oxytocin improves behavioural and neural deficits in inferring others' social emotions in autism. Brain 137, 3073-3086. doi: 10.1093/brain/awu231.

Auyeung, B., Lombardo, M.V., Heinrichs, M., Chakrabarti, B., Sule, A., Deakin, J.B., Bethlehem, R.A., Dickens, L., Mooney, N., Sipple, J.A., Thiemann, P., and Baron-Cohen, S. (2015). Oxytocin increases eye contact during a real-time, naturalistic social interaction in males with and without autism. Transl Psychiatry 5, e507.

Bakermans-Kranenburg, M.J., and van Ijzendoorn, M. H. (2014). A sociability gene? Meta-analysis of oxytocin receptor genotype effects in humans. Psychiatric Genetics, 24(2), 45–51. http://doi.org/10.1097/YPG.0b013e3283643684

Bakermans-Kranenburg, M.J., and Van Ijzendoorn, M.H. (2013). Sniffing around oxytocin: review and meta-analyses of trials in healthy and clinical groups with implications for pharmacotherapy. Translational Psychiatry. 3.

Bartz, J., Simeon, D., Hamilton, H., Kim, S., Crystal, S., Braun, A., Vicens, V., and Hollander, E. (2011). Oxytocin can hinder trust and cooperation in borderline personality disorder. Soc Cogn Affect Neurosci 6, 556-563. doi: 10.1093/scan/nsq085.

Baumgartner T., Heinrichs M., Vonlanthen A., Fischbacher U., Fehr E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron. May 22;58(4):639-50. doi: 10.1016/j.neuron.2008.04.009.

Bell, C.J., Nicholson, H., Mulder, R.T., Luty, S.E., and Joyce, P.R. (2006). Plasma oxytocin levels in depression and their correlation with the temperament dimension of reward dependence. J Psychopharmacol 20, 656-660.

Bethlehem, R.A., Van Honk, J., Auyeung, B., and Baron-Cohen, S. (2013). Oxytocin, brain physiology, and functional connectivity: a review of intranasal oxytocin fMRI studies. Psychoneuroendocrinology 38, 962-974. doi: 10.1016/j.psyneuen.2012.10.

Beery, A.K., Lacey, E.A., and Francis, D.D. (2008). Oxytocin and vasopressin receptor distributions in a solitary and a social species of tuco-tuco (Ctenomys haigi and Ctenomys sociabilis). Journal of Comparative Neurology, 507(6), 1847–1859. http://doi.org/10.1002/cne.21638

Boccia, M.L., Petrusz, P., Suzuki, K., Marson, L., and Pedersen, C.A. (2013). Immunohistochemical localization of oxytocin receptors in human brain. Neuroscience,

253, 155–164. http://doi.org/10.1016/j.neuroscience.2013.08.048

Brookfield, J.F. (2010). Q & A: Promise and pitfalls of genome-wide association studies. BMC Biology, 8(1), 41. http://doi.org/10.1186/1741-7007-8-41

Brosnan, S.F., Parrish, A., Beran, M.J., Flemming, T., Heimbauer, L., Talbot, C.F., Lambeth, S.P., Schapiro, S.J., and Wilson, B.J. (2011). Responses to the Assurance game in monkeys, apes, and humans using equivalent procedures. Proc Natl Acad Sci U S A 108, 3442-3447. doi: 10.1073/pnas.1016269108.

Brosnan S.F., Parrish A., Beran M.J., Flemming T., Heimbauer L., Talbot C.F., Lambeth S.P., Schapiro S.J., Wilson B.J. (2001). Responses to the Assurance game in monkeys, apes, and humans using equivalent procedures. Proc Natl Acad Sci USA. Feb 22;108(8):3442-7. doi: 10.1073/pnas.1016269108. Epub 2011 Feb 7.

Caldwell, H.K., Stephens, S.L., and Young, W.S. (2009). Oxytocin as a natural antipsychotic: a study using oxytocin knockout mice. Mol Psychiatry 14, 190-196. doi: 10.1038/sj.mp.4002150.

Caldwell, H.K., Lee, H.J., Macbeth, A.H., and Young, W.S. (2008). Vasopressin: behavioral roles of an "original" neuropeptide. Progress in Neurobiology, 84(1), 1–24. http://doi.org/10.1016/j.pneurobio.2007.10.007

Cardon L.R., Abecasis G.R. (2003). Using haplotype blocks to map human complex trait loci. Trends Genet. Mar;19(3):135-40.

Cardoso, C., Kingdon, D., and Ellenbogen, M.A. (2014). A meta-analytic review of the impact of intranasal oxytocin administration on cortisol concentrations during laboratory tasks: moderation by method and mental health. Psychoneuroendocrinology 49, 161-170. doi: 10.1016/j.psyneuen.2014.07.

Cesarini, D., Dawes, C. T., Fowler, J. H., Johannesson, M., Lichtenstein, P., and Wallace, B. (2008). Heritability of cooperative behavior in the trust game. Proc Natl Acad Sci USA. 105(10), 3721– 6. http://doi.org/10.1073/pnas.0710069105.

Cesarini, D., Dawes, C. T., Johannesson, M., Lichtenstein, P., and Wallace, B

.(2009). Genetic Variation in Preferences for Giving and Risk Taking. Quarterly Journal of Economics, 12 (4), 809-842. http://hdl.handle.net/1721.1/57904

Coccaro, E.F., Kavoussi, R.J., Hauger, R.L., Cooper, T.B., and Ferris, C.F. (1998). Cerebrospinal fluid vasopressin levels: correlates with aggression and serotonin function in personality- disordered subjects. Arch Gen Psychiatry 55, 708-714.

Chang, W.H., Lee, I.H., Chen, K.C., Chi, M.H., Chiu, N.T., Yao, W.J., Chen, P.S. (2014). Oxytocin receptor gene rs53576 polymorphism modulates oxytocin-dopamine interaction and neuroticism traits–a SPECT study. Psychoneuroendocrinology, 47, 212–220. http://doi.org/10.1016/j.psyneuen.2014.05.020.

Chen, F. S., Kumsta, R., von Dawans, B., Monakhov, M., Ebstein, R.P., and Heinrichs, M. (2011). Common oxytocin receptor gene (OXTR) polymorphism and social support interact to reduce stress in humans. Proc Natl Acad Sci USA., 108(50), 19937–19942. http://doi.org/10.1073/pnas.1113079108

Cherkas L.F., Oelsner E.C., Mak Y.T., Valdes A., Spector T.D. (2004). Genetic influences on female infidelity and number of sexual partners in humans: a linkage and association study of the role of the vasopressin receptor gene (AVPR1A). Twin Res. 2004 Dec;7(6):649-58.

Cottet, M., Albizu, L., Perkovska, S., Jean-Alphonse, F., Rahmeh, R., Orcel, H., Durroux, T. (2010). Past, present and future of vasopressin and oxytocin receptor oligomers, prototypical GPCR models to study dimerization processes. Current Opinion in Pharmacology, 10(1), 59–66. http://doi.org/10.1016/j.coph.2009.10.003

Dadds, M.R., Macdonald, E., Cauchi, A., Williams, K., Levy, F., and Brennan, J. (2014). Nasal oxytocin for social deficits in childhood autism: a randomized controlled trial. J Autism Dev Disord 44, 521-531. doi: 10.1007/s10803-013-1899-3. Dadds M.R., Moul C., Cauchi A., Dobson-Stone C., Hawes D.J., Brennan J., Ebstein R.E. (2013). Methylation of the oxytocin receptor gene and oxytocin blood levels in the development of psychopathy. Dev Psychopathol. 2014 Feb;26(1):33-40. doi: 10.1017/S0954579413000497. Epub 2013 Sep 23.

Devost, D., and Zingg, H.H. (2003). Identification of dimeric and oligomeric complexes of the human oxytocin receptor by co-immunoprecipitation and biolumines-

cence resonance energy transfer. Journal of Molecular Endocrinology, 31(3), 461–471. http://doi.org/10.1677/jme.0.0310461

De Berardis, D., Marini, S., Iasevoli, F., Tomasetti, C., De Bartolomeis, A., Mazza, M., Valchera, A., Fornaro, M., Cavuto, M., Srinivasan, V., Sepede, G., Martinotti, G., and Di Giannantonio, M. (2013). The Role of Intranasal Oxytocin in the Treatment of Patients with Schizophrenia: A Systematic Review. Cns and Neurological Disorders-Drug Targets 12, 252-264.

Dempster, E., Burcescu, I., Wigg, K., Kiss, E., Baji, I., Gadoros, J., Tamás, Z., Kennedy, J.L., Vetró, A., Kovacs, M., and Barr, C.L., (2007). Evidence of an association between the vasopressin V1b receptor gene (AVPR1B) and childhood-onset mood disorders. Archives of General Psychiatry, 64 (10), 1189-1195. http://doi:10.1001/archpsyc.64.10.1189.

De Vignemont, F., and Singer, T. (2006). The empathic brain: how, when and why? Trends in Cognitive Sciences, 10(10), 435–41.

De Waal, F. B. M. (2008). Putting the altruism back into altruism: the evolution of empathy. Annual Review of Psychology, 59(1), 279–300.

Domes, G., Heinrichs, M., Kumbier, E., Grossmann, A., Hauenstein, K., and Herpertz, S.C. (2013). Effects of intranasal oxytocin on the neural basis of face processing in autism spectrum disorder. Biol Psychiatry 74, 164-171.

Dufwenberg, M., and Kirchsteiger, G. (2004). A theory of sequential reciprocity. Games and economic behavior, 47 (2), 268-298. http://doi:10.1016/j.geb.2003.06.003

Ebstein, R.P., Knafo, A., Mankuta, D., Chew, S.H., & Lai, P.S. (2012). The contributions of oxytocin and vasopressin pathway genes to human behavior. Hormones and Behavior, 61(3), 359–79. http://doi.org/10.1016/j.yhbeh.2011.12.014

Ebstein, R.P., Israel, S., Chew, S.H., Zhong, S., and Knafo, A. (2010). Genetics of human social behavior. Neuron, 65(6), 831–44.

Fehr, E., and Schmidt, K. (1999). A theory of fairness, competition, and coopera-

tion. The Quarterly Journal of Economics, (August), 817–868.

Feifel, D., and Reza, T. (1999). Oxytocin modulates psychotomimetic-induced deficits in sensorimotor gating. Psychopharmacology (Berl) 141, 93-98.

Feifel, D., Shilling, P.D., and Belcher, A.M. (2012). The effects of oxytocin and its analog, carbetocin, on genetic deficits in sensorimotor gating. Eur Neuropsychopharmacol 22, 374-378. doi: 10.1016/j.euroneuro.2011.09.

Fink, S., Excoffier, L., and Heckel, G. (2006). Mammalian monogamy is not controlled by a single gene. Proceedings of the National Academy of Sciences of the United States of America, 103(29), 10956–10960.

Furman D.J., Chen M.C., Gotlib I.H. (2011). Variant in oxytocin receptor gene is associated with amygdala volume. Psychoneuroendocrinology. Jul;36(6):891-7. doi: 10.1016/j.psyneuen.2010.12.004. Epub 2011 Jan 3.

Freitag, C. M. (2007). The genetics of autistic disorders and its clinical relevance: a review of the literatura. Molecular Psychiatry, 12, 2–22.

Glimcher, P. (2009). Neuroeconomics: Decision making and the brain. First Edit. San Diego, CA: Elsevier Inc.

Golimbet V., Alfimova M., Abramova L., Kaleda V., Gritsenko I. (2014). Arginine vasopressin 1a receptor RS3 promoter microsatellites in schizophrenia: a study of the effect of the "risk" allele on clinical symptoms and facial affect recognition. Psychiatry Res. 2015 Feb 28;225(3):739-40. doi: 10.1016/j.psychres.2014.11.043.

Golimbet, V., Alfimova, M., Abramova, L., Kaleda, V., and Gritsenko, I. (2015). Arginine vasopressin 1a receptor RS3 promoter microsatellites in schizophrenia: a study of the effect of the "risk" allele on clinical symptoms and facial affect recognition. Psychiatry Research, 225(3), 739–740.

Gordon, I., Vander Wyk, B.C., Bennett, R.H., Cordeaux, C., Lucas, M.V., Eilbott, J.A., Zagoory- Sharon, O., Leckman, J.F., Feldman, R., and Pelphrey, K.A. (2013). Oxytocin enhances brain function in children with autism. Proc Natl Acad Sci U S

A 110, 20953-20958. doi: 10.1073/pnas.1312857110.

Gottesman, I. (1991). Schizophrenia genesis: The origins of madness. A series of books in psychology. New York, NY, US: W H Freeman/Times Books/ Henry Holt and Co.

Guastella, A.J., Einfeld, S.L., Gray, K.M., Rinehart, N.J., Tonge, B.J., Lambert, T.J., and Hickie, I.B. (2010a). Intranasal oxytocin improves emotion recognition for youth with autism spectrum disorders. Biol Psychiatry 67, 692-694.

Guastella, A.J., Howard, A.L., Dadds, M.R., Mitchell, P., and Carson, D.S. (2009). A randomized controlled trial of intranasal oxytocin as an adjunct to exposure therapy for social anxiety disorder. Psychoneuroendocrinology 34, 917-923. doi: 10.1016/j.psyneuen.2009.01.

Guastella, A.J., Kenyon, A.R., Alvares, G.A., Carson, D.S., and Hickie, I.B. (2010b). Intranasal arginine vasopressin enhances the encoding of happy and angry faces in humans. Biol Psychiatry 67, 1220-1222. doi: 10.1016/j.biopsych.2010.03.

Hammock, E.A.D., Lim, M.M., Nair, H.P., and Young, L.J. (2005). Association of vasopressin 1a receptor levels with a regulatory microsatellite and behavior. Genes, Brain and Behavior, 4(5), 289–301. http://doi.org/10.1111/j.1601-183X.2005.00119.x

Hammock, E.A.D., and Young, L.J. (2004). Functional microsatellite polymorphism associated with divergent social structure in vole species. Molecular Biology and Evolution, 21(6), 1057– 1063. http://doi.org/10.1093/molbev/msh104

Hammock, E.A.D., and Young, L.J. (2005). Microsatellite instability generates diversity in brain and sociobehavioral traits. Science (New York, N.Y.), 308(5728), 1630–1634. http://doi.org/10.1126/science.1111427

Haram, M., Tesli, M., Bettella, F., Djurovic, S., Andreassen, O. A., and Melle, I. (2015). Association between Genetic Variation in the Oxytocin Receptor Gene and Emotional Withdrawal, but not between Oxytocin Pathway Genes and Diagnosis in Psychotic Disorders. Frontiers in Human Neuroscience, 9, 9.

Hariri A.R., Tessitore A., Mattay V.S., Fera F., Weinberger D.R. (2002). The amygdala response to emotional stimuli: a comparison of faces and scenes. Neuroimage. 2002 Sep;17(1):317-23.

Hariri A.R., Mattay V.S., Tessitore A., Fera F., Smith WG., Weinberger DR. (2002). Dextroamphetamine modulates the response of the human amygdala. Neuropsychopharmacology. 2002 Dec;27(6):1036-40.

Hariri A.R., Weinberger D.R. (2003). Functional neuroimaging of genetic variation in serotonergic neurotransmission. Genes Brain Behav. Dec;2(6):341-9.

Harrison, P. J., and Law, A. J.(2006). Neuregulin 1 and schizophrenia: genetics, gene expression, and neurobiology. Biological psychiatry. 60(2), 132- 140.

Hurlemann R., Patin A., Onur OA., Cohen MX., Baumgartner T., Metzler S., Dziobek I., Gallinat J., Wagner M., Maier W., Kendrick KM. Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. J Neurosci. Apr 7;30(14):4999- 5007. doi: 10.1523/JNEUROSCI.5538-09.2010.

Insel, T.R. (2010). The challenge of translation in social neuroscience: a review of oxytocin, vasopressin, and affiliative behavior. Neuron, 65(6), 768–79.

Inoue, T., Kimura, T., Azuma, C., Inazawa, J., Takemura, M., Kikuchi, T., Kubota, Y., Ogita, K., and Sagi, F. (1994). Structural organization of the human oxytocin receptor gene. Journal of Biological Chemistry. 269 (51), 32451-32456.

Inoue H., Yamasue H., Tochigi M., Takei K., Suga M., Abe O., Yamada H., Rogers M.A., Aoki S., Sasaki T., Kasai K. (2010).Effect of tryptophan hydroxylase-2 gene variants on amygdalar and hippocampal volumes. Brain Res. 2010 May 17;1331:51-7. doi: 10.1016/j.brainres.2010.03.057.

Israel, S., Lerer, E., Shalev, I., Uzefovsky, F., Riebold, M., Laiba, E., Ebstein, R.P. (2009). The oxytocin receptor (OXTR) contributes to prosocial fund allocations in the dictator game and the social value orientations task. PloS One, 4(5), e5535. http://doi.org/10.1371/journal.pone.0005535

Jacob, S., Brune, C.W., Carter, C.S., Leventhal, B.L., Lord, C., and Cook, E.H.J. (2007). Association of the oxytocin receptor gene (OXTR) in Caucasian children and adolescents with autism. Neuroscience Letters, 417(1), 6–9.

Johnson H.G., Ekman P., Friesen W. (1976). A behavioral phenotype in the de Lange syndrome. Pediatr Res. 1976 Oct;10(10):843-50.

Joiner, T., and Coyne, J.C. (1999). The interactional nature of depression: Advances in interpersonal approaches. (T. Joiner and J. C. Coyne, Eds.). Washington, DC, US: American Psychological Association. http://doi.org/10.1037/10311-000

Kelemenova, S., Schmidtova, E., Ficek, A., Celec, P., Kubranska, A., and Ostatnikova, D. (2010). Polymorphisms of candidate genes in Slovak autistic patients. Psychiatric Genetics, 20(4),137-139. http://doi.org/10.1097/YPG.0b013e32833a1eb3

Kim, S.J., Young, L.J., Gonen, D., Veenstra-VanderWeele, J., Courchesne R., Courchesne, E., Lord, C., Leventhal B.L., Cook, E.H.Jr., and Insel, T.R. (2002). Transmission disequilibrium testing of arginine vasopressin receptor 1A (AVPR1A) polymorphisms in autism. Molecular Psychiatry, 7, 503–507.

Knafo, A., Israel, S., Darvasi A., Bachner-Melman, R., Uzefovsky, F., Cohen, L., Feldman, E., Lerer, E., Laiba, E., Raz, Y., Nemanov, L., Gritsenko, I., Dina, C., Agam, G., Dean, B., Bornstein, G., and Ebstein, R. P. (2008). Individual differences in allocation of funds in the dictator game associated with length of the arginine vasopressin 1a receptor RS3 promoter region and correlation between RS3 length and hippocampal mRNA. Genes, Brain and Behavior, 7, 266–275. doi: 10.1111/j.1601-183X.2007.00341.x

Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., and Fehr, E. (2005). Oxytocin increases trust in humans. Nature, 435(7042):673–6.

Kreps, D.M., Milgrom, P., Roberts, J., and Wilson, R. (1982). Rational Cooperation in the Finitely Repeated Prisoners-Dilemma. Journal of Economic Theory 27, 245-252. doi:10.1016/0022- 0531(82)90029-1.

Krueger, F., Parasuraman, R., Iyengar, V., Thornburg, M., Weel, J., Lin, M., Lipsky,

R.H. (2012). Oxytocin receptor genetic variation promotes human trust behavior. Frontiers in Human Neuroscience, 6(February), 4.

Kuehn, B.M. (2011). Scientists probe oxytocin therapy for social deficits in autism, schizophrenia. JAMA 305, 659-661. doi: 10.1001/jama.2011.117.

Labuschagne, I., Phan, K.L., Wood, A., Angstadt, M., Chua, P., Heinrichs, M., Stout, J.C., and Nathan, P.J. (2010). Oxytocin attenuates amygdala reactivity to fear in generalized social anxiety disorder. Neuropsychopharmacology 35, 2403-2413. doi: 10.1038/npp.2010.123.

Labuschagne, I., Phan, K.L., Wood, A., Angstadt, M., Chua, P., Heinrichs, M., Stout, J.C., and Nathan, P.J. (2011). Medial frontal hyperactivity to sad faces in generalized social anxiety disorder and modulation by oxytocin. Int J Neuropsychopharmacol, 1-14. doi: 10.1017/S1461145711001489.

Landgraf, R., and Neumann, I.D. (2004). Vasopressin and oxytocin release within the brain: a dynamic concept of multiple and variable modes of neuropeptide communication. Frontiers in Neuroendocrinology, 25(3-4), 150–76.

Lang S., Kroll A., Lipinski S.J., Wessa M., Ridder S., Christmann C., Schad L.R., Flor H. (2009). Context conditioning and extinction in humans: differential contribution of the hippocam pus,amygdala and prefrontal cortex. Eur J Neurosci. Feb;29(4):823-32. doi: 10.1111/j.1460-9568.2009.06624.x.

Leng, G., and Ludwig, M. (2008). Neurotransmitters and peptides: whispered secrets and public announcements. The Journal of Physiology, 586(Pt 23), 5625–32. doi.org/10.1113/jphysiol.2008.159103

Levin, R., Heresco-Levy, U., Bachner-Melman, R., Israel, S., Shalev, I., and Ebstein, R.P. (2015). Association between arginine vasopressin 1a receptor (AVPR1a) promoter region polymorphisms and prepulse inhibition. Psychoneuroendocrinology, 34(6), 901–908. http://doi.org/10.1016/j.psyneuen.2008.12.014

Lewis, C.M., Levinson, D.F., Wise, L.H., DeLisi, L.E., Straub, R.E., Hovatta, I., Helgason, T. (2003). Genome scan meta-analysis of schizophrenia and bipolar dis-

order, part II: Schizophrenia. American Journal of Human Genetics, 73(1), 34–48. http://doi.org/10.1086/376549

Loup, F., Tribollet, E., Dubois-Dauphin, M., and Dreifuss, J. J. (1991). Localization of high-affinity binding sites for oxytocin and vasopressin in the human brain. An autoradiographic study. Brain Research, 555(2), 220–32.

Lucas-Thompson, R.G., and Holman, E.A. (2013). Environmental stress, oxytocin receptor gene (OXTR) polymorphism, and mental health following collective stress. Hormones and Behavior, 63(4), 615–624.

Ludwig, M., and Leng, G. (2006). Dendritic peptide release and peptide-dependent behaviours. Nature Reviews. Neuroscience, 7(2), 126–36.

Luo, S., Li, B., Ma, Y., Zhang, W., Rao, Y., and Han, S. (2015). Oxytocin receptor gene and racial ingroup bias in empathy-related brain activity. NeuroImage, 110, 22–31. http://doi.org/10.1016/j.neuroimage.2015.01.042

Luo, S., Ma, Y., Liu, Y., Li, B., Wang, C., Shi, Z., Han, S. (2015). Interaction between oxytocin receptor polymorphism and interdependent culture values on human empathy. Social Cognitive and Affective Neuroscience.

Luppino, D., Moul, C., Hawes, D.J., Brennan, J., Dadds, M.R. (2014). Association between a polymorphism of the vasopressin 1B receptor gene and aggression in children. Psychiatric Genetics, 24(5), 185-190. doi:10.1097/YPG.0000000000000036

MacDonald, K., and Feifel, D. (2012). Oxytocin in schizophrenia: a review of evidence for its therapeutic effects. Acta Neuropsychiatr 24, 130-146. doi: 10.1111/j.1601-5215.2011.

MacDonald, E., Dadds, M. R., Brennan, J. L., Williams, K., Levy, F., and Cauchi, A. J. (2011). A review of safety, side-effects and subjective reactions to intranasal oxytocin in human research. Psychoneuroendocrinology, 36(8), 1114–26.

MacDonald, K., and MacDonald, T. M. (2010). The Peptide That Binds: A Systematic Review of Oxytocin and its Prosocial Effects in Humans. Harvard Review

of Psychiatry, 18(1), 1–21. http://doi.org/10.3109/10673220903523615

Macdonald K., Feifel D. (2012). Oxytocin in schizophrenia: a review of evidence for its therapeutic effects. Acta Neuropsychiatr. Jun 1;24(3):130-146. Epub 2011 Dec 19.

McQuaid, R. J., McInnis, O. A., Matheson, K., and Anisman, H. (2015). Distress of ostracism: oxytocin receptor gene polymorphism confers sensitivity to social exclusion. Social Cognitive and Affective Neuroscience.

McQuaid, R. J., McInnis, O. A., Stead, J. D., Matheson, K., and Anisman, H. (2013). A paradoxical association of an oxytocin receptor gene polymorphism: early-life adversity and vulnerability to depression. Frontiers in Neuroscience, 7, 128. http://doi.org/10.3389/fnins.2013.00128

Mendlewicz, J., Crisafulli, C., Calati, R., Kocabas, N., Massat, I., Linotte, S., ... Serretti, A. (2012). Influence of COX-2 and OXTR polymorphisms on treatment outcome in treatment resistant depression. Neurosci Lett., 10(516(1)), 85–8. http://doi.org/doi: 10.1016/j.neulet.2012.03.063.

Meyer-Lindenberg, A., Domes, G., Kirsch, P., & Heinrichs. (2011). Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. Nature Reviews Neuroscience, 12(9), 524-538. http://doi:10.1038/nrn3044

Meyer-Lindenberg, A, Kolachana, B., Gold, B., Olsh, A., Nicodemus, K.K., Mattay, V., Weinberger, D.R. (2009). Genetic variants in AVPR1A linked to autism predict amygdala activation and personality traits in healthy humans. Molecular Psychiatry, 14(10), 968–975. http://doi.org/10.1038/mp.2008.54

Moons, W.G., Way, B.M., and Taylor, S.E. (2014). Oxytocin and vasopressin receptor polymorphisms interact with circulating neuropeptides to predict human emotional reactions to stress. Emotion (Washington, D.C.), 14(3), 562–572.

Nakajima, M., Görlich, A., & Heintz, N. (2014). Oxytocin Modulates Female Sociosexual Behavior through a Specific Class of Prefrontal Cortical Interneurons. Cell, 295–305. http://doi.org/10.1016/j.cell.2014.09.020.

Neumann, I.D., and Landgraf, R. (2012). Balance of brain oxytocin and vasopressin: implications for anxiety, depression, and social behaviors. Trends in Neurosciences, (Figure 1), 1–11. http://doi.org/10.1016/j.tins.2012.08.004.

Nyffeler J., Walitza S., Bobrowski E., Gundelfinger R., Grünblatt E. (2014). Association study in siblings and case-controls of serotonin- and oxytocin-related genes with high functioning autism. J Mol Psychiatry. Jan 24;2(1):1. doi: 10.1186/2049-9256-2-1.

Ozsoy, S., Esel, E., and Kula, M. (2009). Serum oxytocin levels in patients with depression and the effects of gender and antidepressant treatment. Psychiatry Res 169, 249-252. doi: 10.1016/j.psychres.2008.06.

Pitman, R.K., Orr, S.P., and Lasko, N.B. (1993). Effects of intranasal vasopressin and oxytocin on physiologic responding during personal combat imagery in Vietnam veterans with posttraumatic stress disorder. Psychiatry Res 48, 107-117.

Plomin, R., Defries, J. C., and Knopik, V.S. (2013). Behavioral Genetics.

Powell, S.B., Zhou, X., and Geyer, M.A. (2009). Prepulse inhibition and genetic mouse models of schizophrenia. Brain, 204(2), 282–294.

Rabin, M. (1993). Incorporating fairness into game theory and economics. The American economic review, 83 (5), 1281-1302. http://www.jstor.org/stable/2117561

Radke, S., and De Bruijn, R.A. (2012). The other side of the coin: oxytocin decreases the adherence to fairness norms. Frontiers in human neuroscience, 6, 193. http://10.3389/fnhum.2012.00193

Ratajczak, P., Woźniak, A., and Nowakowska, E. (2013). Animal models of schizophrenia: Developmental preparation in rats. Acta Neurobiologiae Experimentalis, 73(4), 472–484.

Rodrigues, S.M., Saslow, L.R., Garcia, N., John, O.P., and Keltner, D. (2009).

Oxytocin receptor genetic variation relates to empathy and stress reactivity in humans. Proceedings of the National Academy of Sciences, 106(50), 21437–21441. http://doi.org/10.1073/pnas.0909579106

Sala, M., Braida, D., Lentini, D., Busnelli, M., Bulgheroni, E., Capurro, V., Chini, B. (2015). Pharmacologic Rescue of Impaired Cognitive Flexibility, Social Deficits, Increased Aggression, and Seizure Susceptibility in Oxytocin Receptor Null Mice: A Neurobehavioral Model of Autism. Biological Psychiatry, 69(9), 875–882.

Sauer, C., Montag, C., Reuter, M., and Kirsch, P. (2013). Imaging oxytocin x dopamine interactions: an epistasis effect of CD38 and COMT gene variants influences the impact of oxytocin on amygdala activation to social stimuli. Frontiers in Neuroscience, 7, 45. http://doi.org/10.3389/fnins.2013.00045

Seibold, A., Brabet, P., Rosenthal, W., and Birnbaumer, M. (1992). Structure and Chromosomal Localization of the Human Antidiuretic Hormone Receptor Gene. American journal of human genetics. 51(5), 1078- 1083.

Shamay-Tsoory, S.G., Aharon-Peretz, J., and Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. Brain, 132(3), 617–627. http://doi.org/10.1093/brain/awn279

Spanakis, E., Milord, E., & Gragnoli, C. (2008) . AVPR2 variants and mutations in nephrogenic diabetes insipidus: review and missense mutation significance. Journal of cellular physiology. 217 (3), 605-617. http:// doi: 10.1002/jcp.21552

Stoop, R. (2012). Neuromodulation by oxytocin and vasopressin. Neuron, 76(1), 142–59. http://doi.org/10.1016/j.neuron.2012.09.025

Sugimoto, T., Saito, M., Mochizuki, S., Watanabe, Y., Hashimoto, S., and Kawashima, H. (1994). Molecular cloning and functional expression of a cDNA encoding the human v1b vasopressin receptor. The Journal of Biological Chemistry, 269(43), 27088-27092.

Tansey, K.E., Hill, M.J., Cochrane, L.E., Gill, M., Anney, R.J., and Gallagher, L.

(2011). Functionality of promoter microsatellites of arginine vasopressin receptor 1A (AVPR1A): implications for autism. Molecular Autism, 2(1), 3.

Tachibana, M., Kagitani-Shimono, K., Mohri, I., Yamamoto, T., Sanefuji, W., Nakamura, A., Oishi, M., Kimura, T., Onaka, T., Ozono, K., and Taniike, M. (2013a). Long-term administration of intranasal oxytocin is a safe and promising therapy for early adolescent boys with autism spectrum disorders. J Child Adolesc Psychopharmacol 23, 123-127. doi: 10.1089/cap.2012.0048.

Tachibana, M., Kagitani-Shimono, K., Mohri, I., Yamamoto, T., Sanefuji, W., Nakamura, A., Oishi, M., Kimura, T., Onaka, T., Ozono, K., and Taniike, M. (2013b). Long-Term Administration of Intranasal Oxytocin Is a Safe and Promising Therapy for Early Adolescent Boys with Autism Spectrum Disorders. Journal of Child and Adolescent Psychopharmacology 23, 123-127. doi:10.1089/Cap.2012.0048.

Thibonnier, M. (1992). Signal transduction of V1-vascular vasopressin receptors. Regulatory Peptides, 38(1), 1–11.

Thibonnier, M., Graves, M. K., Wagner, M. S., Chatelain, N., Soubrier, F., Corvol, P., Jeunemaitre, X. (2000). Study of V(1)-vascular vasopressin receptor gene microsatellite polymorphisms in human essential hypertension. Journal of Molecular and Cellular Cardiology, 32(4), 557–564. http://doi.org/10.1006/jmcc.2000.1108

Thibonnier, M., Graves, M.K., Wagner, M.S., Auzan, C., Clauser, E., and Willard, H.F. (1996). Structure, Sequence, Expression, and Chromosomal Localization of the Human V1a Vasopressin Receptor Gene. Genomics, 31, 327 –334.

Thibonnier, M., and Schork, N.J. (1995). The genetics of hypertension. Current Opinion in Genetics and Development, 5(3), 362–370.

Thompson, R.J., Parker, K.J., Hallmayer, J.F., Waugh, C.E., and Gotlib, I.H. (2011). Oxytocin receptor gene polymorphism (rs2254298) interacts with familial risk for psychopathology to predict symptoms of depression and anxiety in adolescent girls. Psychoneuroendocrinology, 36(1), 144–147.

Thompson, S.M., Hammen, C., Starr, L.R., and Najman, J.M. (2014). Oxytocin

receptor gene polymorphism (rs53576) moderates the intergenerational transmission of depression. Psychoneuroendocrinology, 43, 11–19.

Trikalinos, T.A., Karvouni, A., Zintzaras, E., Ylisaukko-oja, T., Peltonen, L., Järvelä, I., and Ioannidis, J.P.A. (2006). A heterogeneity-based genome search meta-analysis for autism- spectrum disorders. Molecular Psychiatry , 11, 29–36.

Turnpenny, Peter (2004). "Emery's Elements of Medical Genetics" (PDF) (12th ed.). Elsevier.

Van West, D., Del-Favero, J., Aulchenko, Y., Oswald, P., Souery, D., Forsgren, T., Claes, S. (2004). A major SNP haplotype of the arginine vasopressin 1B receptor protects against recurrent major depression. Molecular Psychiatry, 9(3), 287–292. http://doi.org/10.1038/sj.mp.4001420

Veening, J.G., and Olivier, B. (2013). Intranasal administration of oxytocin: behavioral and clinical effects, a review. Neurosci Biobehav Rev 37, 1445-1465. doi: 10.1016/j.neubiorev.2013.04.

Wassink, T.H., Brzustowicz, L.M., Bartlett, C.W., and Szatmari, P. (2004). The search for autism disease genes. Mental Retardation and Developmental Disabilities Research Reviews, 10(4), 272-283. http://doi:10.1002/mrdd.20041.

Wang J., Qin W., Liu B., Zhou Y., Wang D., Zhang Y., Jiang T., Yu C. (2014). Neural mechanisms of oxytocin receptor gene mediating anxiety-related temperament. Brain Struct Funct. Sep;219(5):1543-54. doi: 10.1007/s00429-013-0584-9.

Williams, K.D., Cheung, C.K.T., and, Choi, W. (2000). Cyberostracism: Effects of being ignored over the Internet. Journal of Personality and Social Psychology, 79(5), 748- 762.http://dx.doi.org/10.1037/0022-3514.79.5.748

Wu, N., Li, Z., & Su, Y. (2012,). The association between oxytocin receptor gene polymorphism (OXTR) and trait empathy. Journal of Affective Disorders, 138 (3), 468-472. doi:10.1016/j.jad.2012.01.009.

Wu, S., Jia, M., Ruan, Y., Liu, J., Guo, Y., Shuang, M., Zhang, D. (2005). Positive

association of the oxytocin receptor gene (OXTR) with autism in the Chinese Han population. Biological Psychiatry, 58(1), 74–77.

Yan, Y., Wang, Y.L., Su, Z., Zhang, Y., Guo, S.X., Liu, A.J., Yang, J. (2015). Effect of oxytocin on the behavioral activity in the behavioral despair depression rat model. Neuropeptides, 48(2), 83– 89. http://doi.org/10.1016/j.npep.2014.01.001

Yirmiya, N., Rosenberg, C., Levi, S., Salomon, S., Shulman, C., Nemanov L., Dina C., and Ebstein, R.P.(2006) . Association between the arginine vasopressin 1a receptor (AVPR1a) gene and autism in a family-based study: mediation by socialization skills. Molecular Psychiatry, 11, 488–494. http:// doi:10.1038/sj.mp.4001812

Young, L., Nilsen, R., Waymire, K., and MacGregor, G. (1999). Increased affillative response to vasopressin in mice expressing the V1a receptor from a monogamous vole. Nature, 400(August), 1998–2000.

Zai, C.C., Muir, K.E., Nowrouzi, B., Shaikh, S.A., Choi, E., Berall, L., Kennedy, J.L. (2012). Possible genetic association between vasopressin receptor 1B and child aggression. Psychiatry Research, 200(2-3), 784–788.

Zak P.J., Stanton A.A., Ahmadi S. (2007). Oxytocin increases generosity in humans. PLoS One. 2007 Nov 7;2(11):e1128.

Zink, C.F., and Meyer-Lindenberg, A. (2012). Human neuroimaging of oxytocin and vasopressin in social cognition. Hormones and Behavior, 61(3), 400–409.

4 Unmasking reciprocity: Late frontal negativity dicriminates outcomes and intentions in trust-repayment behaviour

Paper submitted to *Frontiers in Social Psychology* by:

Mauricio Aspé-Sánchez[1,2,3,4,*], Paola Mengotti[4,5], Raffaella Rumiati[4], Carlos Rodríguez-Sickert[2], John Ewer[3] & Pablo Billeke[1,2]

[1] División de Neurociencia (NeuroCICS), Centro de Investigación en Complejidad Social (CICS), Facultad de Gobierno, Universidad del Desarrollo, Chile.

[2] Centro de Investigación en Complejidad Social (CICS), Facultad de Gobierno, Universidad del Desarrollo, Chile.

[3] Instituto de Neurociencia, Universidad de Valparaíso, Valparaíso, Chile.

[4] Neuroscience Area, Scuola Internazionale di Studi Avanzati (SISSA), Trieste, Italy.

[5] Cognitive Neuroscience, Institute of Neuroscience & Medicine (INM-3), Research Centre Juelich, Germany.

[*] Address for correspondence: Mauricio Aspé-Sánchez, Centro de Investigación en Complejidad Social, Facultad de Gobierno, Universidad del Desarrollo, Av. Las Condes 12461, Of. 307, Las Condes, Santiago, Chile. Email: mauricio.aspe.s@gmail.com

Running title: ERP negativity in repayment behaviour.

4.1 Abstract

Altruism (a costly action that benefit others) and reciprocity (the repayment of acts in kind) differ in that the former expresses preferences about the outcome of a social interaction, whereas the latter requires, in addition, ascribing intentions to others. Interestingly, individuals behaviour and neurophysiological activity under outcomes-versus intention-based interactions have not been compared using different endowments, in the same subject and during the same session. Here, we used a mixed version of the Dictator and the Investment games, together with electroencephalography, to uncover a subject's behaviour and brain activity when challenged with endowments of different sizes in contexts that call for an altruistic (outcome-based) versus a reciprocal (intention-based) response. We found that subjects displayed positive or negative reciprocity (reciprocal responses greater or smaller than that for altruism, respectively) depending on the amount of trust they received. Furthermore, a subject's late frontal negativity differed between conditions, predicting responses to trust in intentions-based trials. Finally, structures related with mentalising and cognitive control were the cortical sources of this activity. Thus, our work disentangles the behavioural components present in trust-repayment behaviour, and sheds light on the neural activity underlying the integration of outcomes and perceived intentions in human economic interactions.

Keywords: Altruism, medial prefrontal cortex, event-related potentials, positive and negative reciprocity, temporoparietal junction.

4.2 Introduction

Homo sapiens display a large number of cooperative behaviours (Andreoni and Miller, 2003; Axelrod and Hamilton, 1981; Berg et al., 1995; Forsythe et al., 1994; Bolton and Ockenfels, 2000; Camerer, 2003; Fehr and Schmidt, 2001; Johnson and Mislim, 2001). During cooperation, individuals usually have to trust others, thus the trustors risk being betrayed by defection from their counterpart (Berg et al., 1995; Coleman, 2000; Camerer, 2003). This implies that cooperation depends critically on other-regarding dispositions, such as reciprocity and altruism (Nowak and Sigmund, 2005, Fehr et al., 2001, Camerer, 2003). Yet, the interplay between reciprocity and altruism in trust-repayment behaviour is still not well understood, neither at the behavioural nor at the neurobiological level.

Altruism is defined as a costly, unconditional act, that benefits another individual (Wilson, 1975; Trivers, 1985; Hamilton, 1996). Being an "outcome-based" behaviour (Cox, 2004, Ashraf et al., 2006), altruism does not necessarily require ascribing cooperative intentions to others. Altruism is usually measured using the Dictator game (DG; Forsythe et al., 1994). In a DG, a subject, the "dictator", receives an amount of money, and has the choice of showing altruism by donating some part of this money to another subject, the "recipient". Reciprocity, on the other hand, is defined as the tendency to return acts in kind (Nowak and Sigmund, 2005). Positive reciprocity involves the costly rewarding of an act to which we ascribe positive intentions, and negative reciprocity the costly sanctioning of an act to which we ascribe negative intentions (Gouldner, 1960). Therefore, it constitutes an "intention-based" behaviour (Berg et al., 1995; Cox, 2004, Ashraf et al., 2006). The Investment game (IG) is a fruitful interaction structure to study the effects of reciprocal preferences. In this game (and in its binary decision version, the Trust game; TG; McCabe et al.,

2003), the experimenter gives an amount of money to a subject, called "the trustor", who can then trust part of this money to another subject, called "the trustee". The trustee, in turn, has the opportunity to show reciprocity by sending back a certain amount to the trustor (Berg et al., 1995). The amount repaid by the trustee has classically been considered (positive) reciprocity; however, the trustee's behaviour might not reflect the cooperative intentions of the trustor – as required if this interaction did indeed correspond to reciprocity, but might reflect the outcome-based concerns of the trustee, hence only representing altruism. Consequently, measuring altruism in the trustee would reveal how much of the trust-repayment behaviour is in fact positive reciprocity: alternatively, it might represent an unconditional, outcome-based behaviour, or even negative reciprocity (Camerer, 2003; Cox, 2004; Ashraf et al., 2006), if the amount sent in an intention-based condition (i.e., in the IG) is less than the amount sent in an outcome-based condition (i.e., in the DG).

Insights from neurocognitive studies have not disentangled intention- from outcome-based behaviours. Reports using the TG have shown that the mentalising system, mainly the temporoparietal junction (TPJ; Decety and Lamm, 2007; Frith and Frith, 1999), is activated when a trustee reciprocates a trustor's risky allocation (van den Bos et al., 2009). The TPJ has also been shown to be important in the control of selfish impulses (Hutcherson et al., 2015). The cognitive control system, which is crucial for the inhibition of selfishness, and for strategic and normative decision making, may also be involved since the anterior cingulate (ACC; Delgado et al., 2005; Shenhav et al., 2013; van den Bos et al., 2009) and the dorsolateral prefrontal cortices (DLPFC; Baumgartner et al., 2011; Chang et al., 2011; Yamagishi et al., 2016) show increased activity when trustees repay trust with an amount that is smaller or greater than what they think the trustor expects to be repaid, respectively (Chang

et al., 2011). However, both the mentalising and the cognitive control networks are also involved in outcome-based conditions: The activity of the right TPJ correlates with how subjects value the outcomes of others (Hutcherson et al., 2015), whereas the connectivity between the ACC and the anterior insula predicts empathy-driven (outcomes-based) versus reciprocity-driven (intentions-based) altruism (Hein et al., 2016). Electrophysiologically, the frontomedial negativity (FMN), a family of event-related potential (ERP) deflections classically related to performance monitoring (Holroyd and Coles, 2002) (and whose source appears to be the medial prefrontal cortex; Holroyd and Coles, 2002; Bileke et al., 2013; Ullsperger et al., 2014; Cavanagh and Frank, 2014), is more pronounced when subjects receive an unfair allocation from a friend playing the role of dictator (Wu et al., 2011). Importantly, to date no studies have used the greater temporal resolution of EEG to disentangle intention- and outcome-based neural activity.

Here we used a "Dictator/Investment" game (DIG), which mixed the canonical DG and IG (based on Cox, 2004; see also Ashraf et al., 2005; see Figure 8), to compare subjects' ERPs activity (i) when they received an amount from a human trustor (IG conditions), which signaled that they are under an intentions-based scenario, versus (ii) when they received the same amount from a computer (DG conditions), which signaled that they are under an outcomes-based scenario (see Figure 8 and Methods). In addition, we compared the subsequent allocation that subjects made under both scenarios. In summary, our work disentangles the different behavioural components underlying the repayment of trust, and implicates brain networks involved in mentalising and cognitive control in the process of integrating outcomes and perceived intentions when humans engage in economic interactions.

4.3 Methods

4.3.1 Participants

Twenty right-handed undergraduate students (mean age = 21.2 years; s.d. = 2.07 years; min = 18 years; max = 25 years; 45% women) participated in the experiment. Participants were instructed to abstain from exercise, caffeine, and alcohol, starting the night before the sessions. Subjects with chronic diseases, mental disorders, medication, or those who abused drugs, were excluded. All subjects approved and signed a written informed consent form. Then, they read written and listened to verbal instructions explaining the task. Finally, they answered a 7 item questionnaire, to ensure that they had understood the logic of the game. All participants answered correctly all questions. All experimental sessions were carried out in the EEG Lab of the Neuroscience Area of the Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy. The experiments were performed according to the Declaration of Helsinki, and approved by the SISSA bioethics committee.

4.3.2 Instruments

The DIG (Dictator-Investments Game). The DIG, which we introduce here, combines the classical IG (Berg et al., 1995) and DG (Forsythe et al., 1994) setups, based on the experiments made by Cox (2004) (see also Ashraf et al., 2006). We call P_1 the trustor and P_2 the trustee (as in a classical IG). In our experiment the focus is on P_2. P_2 subjects performed 60 trials of a recurrent-interactions DIG. In each trial, subjects played either an IG or a DG, which was decided using a pseudo-random distribution programmed so that each player played 30 IG trials and 30 DG trials. In both scenarios, the initial maximum amount of money available was equal to €12, and the exchange factor was equal to 3 (see Figure 8 and above). The DIG trials

were divided into 3 block of 20 trials each. Participants were told that they would play with 20 different P_1s, located in a different and dedicated room. Thus, they played three times with each P_1 (once in each block), always in a random order, which prevented reputation-building motives. After they completed the task, they received an amount equal to the outcome of a random trial, plus €10-base. All P_1's and COM allocation were computer simulations sending a pseudo-random allocation drawn from a uniform distribution in the range of [0, 12] €.

IG trials. In the IG trials of the DIG, P_2s began with €0. P_1s then sent an allocation A_1 in the range [0, 12] €, which was multiplied by 3, and given to P_2. P_2 then decided how much of the amount they received (3 x A_1 €) to send back to the P_1 they were paired with in this trial. The allocation made by P_2 (A_2) is considered to be an intentions-based allocation, given that the behaviour of P_2 should be influenced by the ascription of an intention to trust or cooperate, signaled by the allocation A_1 of the P_1 (see Berg et al., 1995; Cox, 2004, Ashraf et al., 2006).

DG trials. In the DG trials of the DIG, P_2 also began the game with 0 €. Unlike the IG, however, P_2 was told that the amount A_1 (also in the range of [0, 12] €) they received would be decided not by P_1, but a computer (COM). It was then multiplied by 3 and given to P_2. As in the IG, P_2s then decided how much of the amount they received (3 x A_1 €) to send to the P_1 they were paired with in this trial. In this case, however, the allocation made by the P_2 (A_2) is considered to be an outcome-based allocation, given that their behaviour should be influenced not by the ascription of a cooperative intention of P_1 (because P_1 did not decide how much COM send to P_2), but by concerns regarding the distribution of the total amount of money available.

4.3.3 Experimental procedure

Behavioural task. Participants were seated in front of a computer monitor in a soundproof cabin. All electrical devices that could interfere with EEG acquisition were turned off. Stimuli were presented using the "Presentation" software (www.neurobs.com).

In order to ensure that participants understood the dynamics of the game, they first played three practice trials of the DIGl. Then, participants were left alone in the room and began to play the DIG. The description of the flow of the game is detailed in Supplementary Material (see also Figure 8).

Analyses of behavioural data. Behavioural data were analyzed using the R software (R Core Team, 2016). To test if the results were normally distributed, we used the Shapiro-Wilk normality tests (R function shapiro.test). To analyze differences between A_2s in the IG versus the DG conditions, we used a two-sided Mann-Whitney test (R function wilcox.test), with a confidence interval of 95%. For this, we divided the trials into three categories, depending on the amount received by P_2: when $\hat{A}_1$ (i.e., A_1 normalised as $A_1/T = A_1/12$) was less than 1/3, when $\hat{A}_1$ was between 1/3 and 2/3, and when $\hat{A}_1$ was greater than 2/3. We then compared the subjects' behaviour in the IG and DG, separately for each of the three categories, also using a 95% confidence interval.

Regression analysis was performed using a Linear Mixed-Effect Model, with error clusterized by subjects (R function lme). We expressed the trustor's response using the following model:

$$\hat{A}_2(t) = \beta_0 + \beta_1 \hat{A}_1(t) + \beta_2 \hat{A}_1(t) IG(t) + \beta_3 IG(t) + \epsilon \quad (1)$$

where $\hat{A}_2$(t) is the normalised amount subjects sent to P_1 ($A_2/3A_1$; i.e., the amount P_2 sent divided the total amount available for this trial) in trial t, $\hat{A}_1$ the normalised amount P_2 received ($\hat{A}_2/12$), and IG is a dummy variable that was given the value 1 when P_2 was partnered with a "human" (IG trials) and 0 when they were partnered with a COM (DG trials).

4.3.4 EEG acquisition

EEGs were recorded continuously while participants played the DIG. Recordings were made from an array of 128 silver-chloride active electrodes mounted on an elastic cap, using standard positioning (10-20 system). Reference electrodes were placed on the left and right mastoids (A_1/A_2). EEG signals were sampled at 1024 Hz, and amplified using an Active-Two amplifier system (Biosemi, Amsterdam, Netherlands). The ground reference consisted of two separate electrodes: Common Mode Sense (CMS) active electrode and a Driven Right Leg (DRL) passive electrode. Electrode sockets were filled with conducting gel to increase signal quality. Electrode offset was kept below 25 μV. An on-line analog low-pass acquisition filter was set at 256 Hz. Data acquisition was made using the Actiview605-Lores software (www.biosemi.com).

4.3.5 ERP analysis

Offline EEG data analysis was performed using EEGLab (Delorme and Makeig, 2004) and LANToolbox, a Matlab toolbox built using algorithms implemented in Fieldtrip, EEGLab, and Cronix, and specifically designed for advanced EEG signal analyses

(http://neuroCICS.udd.cl/; see, for example, our previous work: Larrain-Valenzuela et al., 2017). EEG data for 3 of the 20 participants were excluded because they had more than 40% of trials rejected, based on semiautomatic criteria.

Preprocessing was made by applying a band-pass filter between 0.1 and 100 Hz to the raw signal. Epochs were extracted in the time range between [-1.5, 1.5] s, centered on the time when subjects were notified about the allocation they received (A_1). Eye-blinks were identified applying a threshold of 100 μV, and removed using independent-component analysis (ICA) on the signal, as implemented in the EEGLab Toolbox. Noisy trials were identified by visual inspection and excluded. Signals were filtered using a low pass filter of 40 Hz, and evoked activity was computed as the average signal recorded at each electrode, for all the participants. Baseline was based on the signal recorded $[-0.3, -0.05]$ s. For visualization purposes, a low-pass filter of 20 Hz was applied.

4.3.6 Source estimation

For the estimation of cortical sources, electrode activity (first referenced to mastoids electrodes) was re-referenced to the average of all electrodes. A brain model taken from the anatomy of a standard human brain was used to projects scalp activity onto the cortical surface (Montreal Neurology Institute; MNI/Colin27). We defined 5000 cortical sources with 3 orthogonal dipoles each (thus, $3X$ sources). A three layer conductivity model (brain, skull, and scalp) and a physical forward model (Clerc et al., 2010) were calculated.

Source estimation was computed using an inverse solution based on a weighted minimum norm estimate (wMNE), based on Billeke et al. (2015), as implemented in Brainstorm software (Tadel et al. 2011). Current source density time series for each

cortical source was computed with unrestrained dipole orientation, for the average for each condition and for each subject. The activity x of N electrodes over time (t), $X(t) = [x_1(t), x_2(t), ..., x_n(t)]$, was assumed to be linearly correlated with a set Y of M cortical sources over time $Y(t) = [y_1(t), y_2(t), ..., y_m(t)]$ and with additive noise $N(t)$: $X(t) = LY(t) + N(t)$, where L is the physical forward model. An inverse solution was derived as $Y(t) = MX(t) = RLT(LRLT + \lambda 2C)–X(t)$, where M is the inverse operator, R is the source covariance, C the noise covariance, and λ a regulatory parameter, set to 1/3 (Lin et al., 2006). With this, we obtained a time-series of the electrical activity for each cortical source.

4.3.7 Statistical analyses of EEG data

For ERP analysis, we took the grand average ERP for all subjects for the time epochs when they were notified about amount A_1, and grouped them depending on whether they were measured under the IG or the DG conditions. We compared, separately, the results obtained under both conditions using Wilcoxon signed rank test, as implemented in the LAN Toolbox. Signals were projected onto a three-dimensional space, and the adjacent areas with significant differences in this space were corrected using a cluster permutation test (Maris and Oostenveld, 2007). We defined the cluster as a groups of adjacent points that showed the same effect, with a threshold of $p < 0.05$. In order to compare the EEG activity obtained in trials when P_2s received a high versus low A_1 feedback, we sorted the results depending on whether A_1 was greater or smaller than €6, then compared the results from both groups for statistical differences, under IG and DG conditions.

4.3.8 EEG activity and behavioural parameters

To investigate the relationship between a subject's EEG activity and the correlation between A_2 and A_1 (β-value in the regression of Equation 1), we calculated, for each subject and separately for the IG and the DG trials, the average centromedial activity, composed of activity from electrodes [C12, C13, C14, C19 (AFz), C20, C21 (Fz), C25, C26, C27]. To assess the relationship between β-values and a subject's frontomedial activity, we specified two separate models (R command lm), one for the IG and the other for the DG trials, both of the form:

$$\beta(s) = \gamma_0 + \gamma_1 FMA(s) + \epsilon \tag{2}$$

where $FMA(s)$ was the subject's average frontomedial activity in the time epoch when they were notified about A_1, and β is the estimated β-value of the regression of A_2 on A_1, calculated for each subject (see Equation 1).

4.4 Results

On average, subjects (P_2s) reciprocated (to P_1s) an allocation Â2 (A_2 / $3A_1$; see Methods) of similar magnitude in the DG and IG conditions of the DIG (average for DG: 0.19; median: 0.17; s.d.: 0.17; average for IG: 0.21; median: 0.18; s.d.: 0.19; p = 0.39, W = 160860; Mann-Whitney test) (Figure 9A). Interestingly, we found a strong correlation between a given subject's behaviour in the DG and their behaviour in the IG ($r = 0.72$; $t = 4.36$; $p = 0.0004$; Pearson correlation) (Figure 9B).

To understand how subjects repay trust under outcome-based versus intention-based

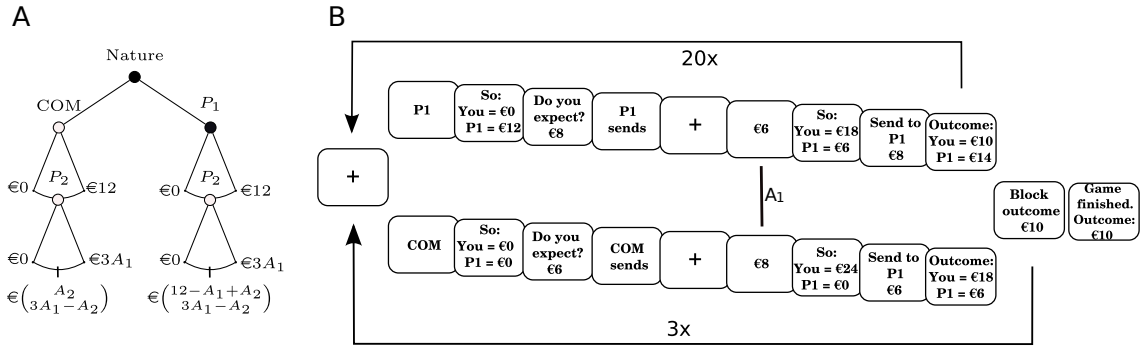


Figure 8: Experimental protocol. A. Behavioral protocol. Schematic of the Dictator/Investment game (DIG) used here. Whether the subject faced a Dictator (DG; open node) or an Investment (IG; closed node) condition was decided randomly with a probability of 0.5 for each. Subjects always played as P_2. The payoff matrix is at the bottom of the tree, with payoffs for P_1 and P_2 shown in the first and second row, respectively. B. Flow of the game. Subjects played a total of 60 trials (in 3 blocks of 20 trials), consisting of 30 trials under the IG conditions (upper flow) and 30 trials under the DG conditions (lower flow). Trigger (vertical line) mark the moment when the subject was notified that an allocation (A_1) had been made.

conditions, we regressed subjects' $\hat{A}_2$ on the $\hat{A}_1$s they received under the DG versus the IG conditions of the DIG (Equation 1). Under the DG condition we found that subjects sent the same proportion of their endowment to P_1, irrespective of the amount P_2s received ($\beta 1 = 0.085$; $p = 0.16$) (Figure 9C and Table 1). In contrast, under the IG condition we obtained a significant and positive β-value ($\beta_2 = 0.23$; $p = 0.0001$). Thus, this dependance of A_2 on A_1 is specific for intention-based behaviours.

We next tested for the existence of regions in which negative and positive reciprocity could be observed. We defined these regions as investment ranges in which subjects playing in the IG condition would send amounts smaller than or greater than they would under the DG condition, respectively (see Methods). Wilcoxon tests revealed significant differences when $\hat{A}_1$ was in the $[0, 1/3]$ range ($W = 6818.5$; $p = 5.16^{-6}$; difference in location $= -4.17^{-5}$) and in the $[2/3, 1]$ range ($W = 35656$, $p = 1.23^{-8}$;

difference in location = 0.083), failing to show a difference in location for $\hat{A}_1$ in $(1/3, 2/3)$ ($W = 18382$, $p = 0.49$). Thus, the DIG setup allowed us to unmask three different behaviours: (i) negative reciprocity, where the amount subjects sent was lower in the IG than in the DG; (ii) an area where the behaviour of subjects between conditions was indistinguishable; and (iii) positive reciprocity, where the amount subjects sent was higher in the IG than in the DG (see Figure 9C-D).

We next analyzed the ERPs of subjects centered on epochs in the range between $[-0.5, 1]$ s relative to when they were notified about the allocation A_1 they received. We found a significant modulation between 560 and 680 ms after stimulus presentation ($p < 0.01$; cluster-based permutation test; cluster threshold detection: $p < 0.05$; Wilcoxon test paired samples) in a centromedial ROI of electrodes (a priori selection, see Methods). Specifically, subjects displayed a more prominent frontal negativity when they were notified about A_1 in the DG versus IG condition (see Figure 10A). Estimations of the cortical sources of these differences projected to the left dorsolateral prefrontal cortex (DLPFC), the left anterior cingulate cortex (ACC), and the left temporoparietal junction (TPJ) ($p < 0.01$; uncorrected; $FDR : q = 0.05$, Figure 10B).

We next focused specifically on the results obtained under outcome-based (DG) condition. As the centromedial negativity was more prominent when subjects were notified about P_1 in the DG trials, we hypothesized that the magnitude of the received endowment might modulate this potential. For this, we divided the trials depending on whether A_1 was above or below the median of the range of A_1 ($A_1 =$ €6). We found significant differences between conditions ($p < 0.01$; cluster-based permutation test; cluster threshold detection: $p < 0.05$; Wilcoxon test paired samples), with ERPs for trials in which subjects received an A_1 ¿ €6 being associated

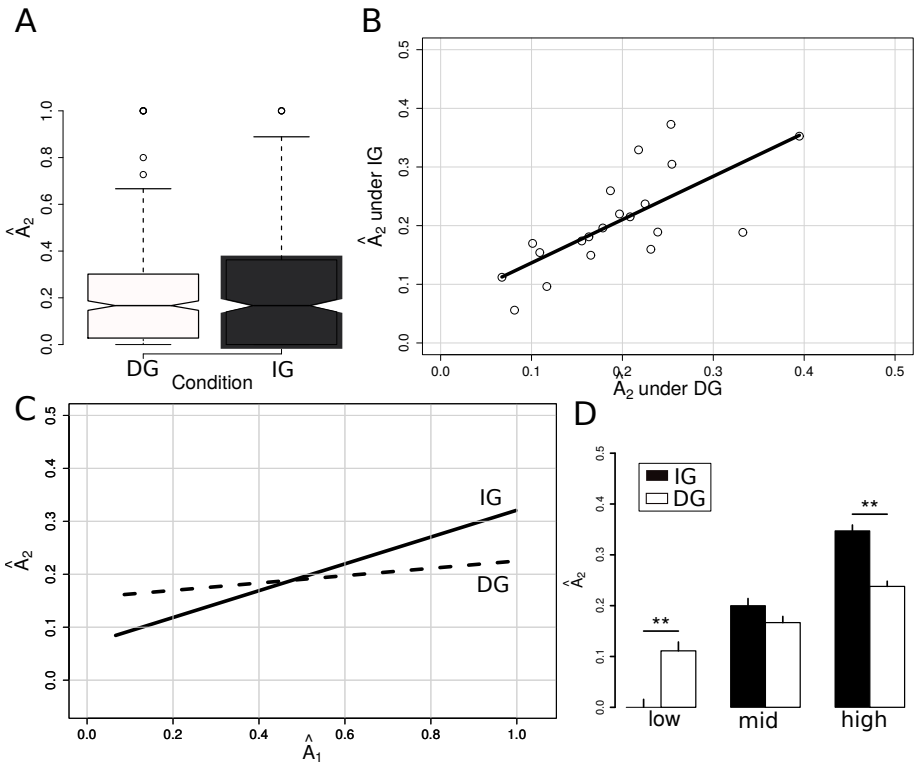


Figure 9: Summary of allocations made in the DG and the IG conditions. A. Barplot of allocations made under DG (white) and IG (black) conditions. No significant differences were found when we considered the allocations made over the entire range of possible endowments. B. Scatterplot of mean individual allocations (open circles) under the DG (X-axis) versus the IG (Y-axis) conditions; black line corresponds to the linear regression ($r = 0.72$; $p = 0.0004$). Areas of positive and negative reciprocity as a function of the amount received. C. Participants' normalised allocations as a function of the normalised amount they received (X-axis). Lines represent the linear regression of A_2 on A_1 (see Equation 1) under IG (solid line) and DG (dashed line) conditions. D. Bar plot showing the mean allocations made by subjects (Y-axis) under the IG (black bars) and DG (white bars) conditions, for different levels of received trust (X-axis); low: $A1 < 1/3$; mid: $1/3 =< A1 < 2/3$; and high: $A1 >= 2/3$. ** indicates $p < 0.01$.

with more negative centromedial activity, as compared to the ERPs for trials in which A_1 ¡ €6. No such differences were found for IG trials (Figure 12A). Thus, the amplitude of the frontal negativity depends on the magnitude of the amount that subjects receive specifically in the outcomes-based conditions. Cortical projection

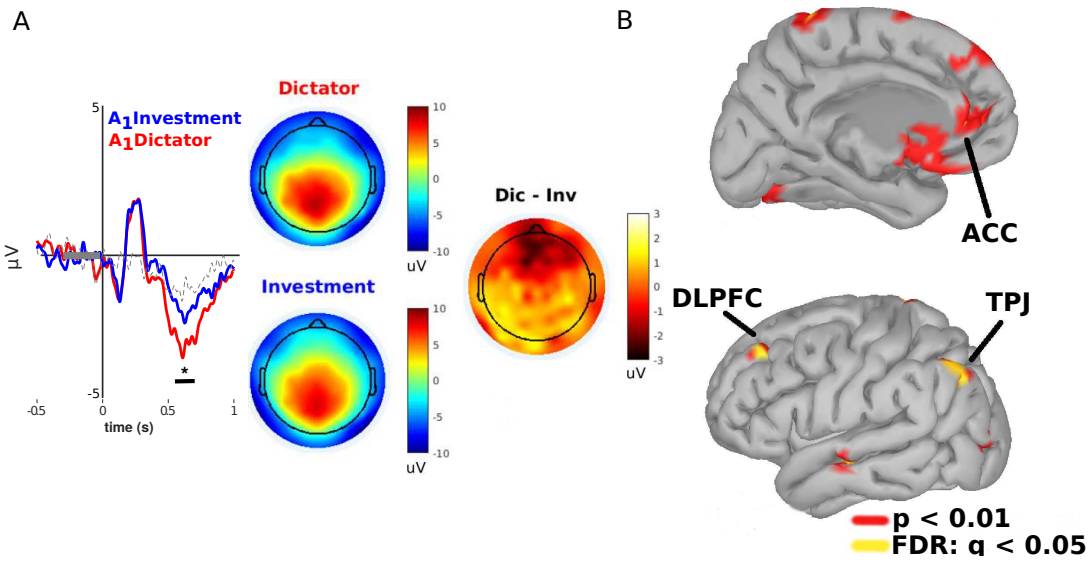


Figure 10: FMN negativities are greater under DG condition than under IG condition when subjects are notified about the allocation they will receive (A1). **A.** Left: ERPs amplitude (Y-axis) of subjects when they received an allocation from a human (blue lines) versus a COM (red lines). Statistically significant differences occurred between 550 and 680 ms (X-axis) after stimulus onset. Right: Scalp potentials distribution. **B.** Cortical source projections ($FDR : q < 0.05$; $p < 0.01$, uncorrected). * indicates $p < 0.05$.

suggests that the right DLPFC was the cortical source of this difference in ERPs ($p < 0.01$; uncorrected; $FDR : q < 0.05$, Figure 12B).

Finally, we focused on the centromedial negativity that subjects displayed in the IG trials. We explored whether the individual's mean potentials in the frontomedial ROI between 560 and 680 ms after the stimulus onset could be predictive of the subject's behaviour, specifically how subjects responded to an additional unit of trust (i.e., the values of β_2 in the regression of $\hat{A}_2$ on $\hat{A}_1$; see Equation 1 and above). Our analyses revealed a significant correlation between both variables ($\gamma_1 = -0.038$; $p = 0.046$; see Equation 2), showing that subjects with more negative values in this frontomedial cluster presented greater β-values in the behavioural regression (Figure 12C), thus predicting how subjects responded to trust. We did not observe this association for

the DG condition ($\gamma_1 = 0.0032$; $p = 0.8$; see Table 2). Cortical source estimations of the coefficients of this regression also projected to the left DLPFC ($p < 0.01$; uncorrected; $FDR : q < 0.05$, Figure 12D).

4.5 Discussion

Here we combined electroencephalography and two canonical behavioural economic games to investigate a subject's behaviour and neurophysiological activity in contexts that called for an altruistic response – which requires only concerns about outcome, versus a reciprocal one – which require concerns about both outcome and intention.

We directly compared these two responses by devising a mixed Dictator-Investment game, and comparing the amount subjects sent in an intention-based condition (the Investments game trials) with the amount that same subject sent back under an outcome-based condition (the Dictator game trials). We found that subjects displayed other-regarding behaviours, allocating amounts greater than €0 most of the time. In addition, a subject's behaviour depended on which condition they were playing. Indeed, we found that only under the intentions-based condition did the proportion of the endowment they sent increase when they received greater allocations, suggesting that the ascription of intention caused the differences in behaviour.

Our results extend those from previous reports, in which subjects' altruism and reciprocity were compared only for specific values of endowment (for outcomes-based conditions) or received trust (for intentions-based conditions) (i.e., Cox, 2004; Ashraf et al., 2006). By testing how the same subject behaved over a whole range of possible endowments, our results allow us to distinguish three different phenomena: (i) the previously reported positive reciprocity (Cox, 2004; Ashraf et al., 2006; McCabe

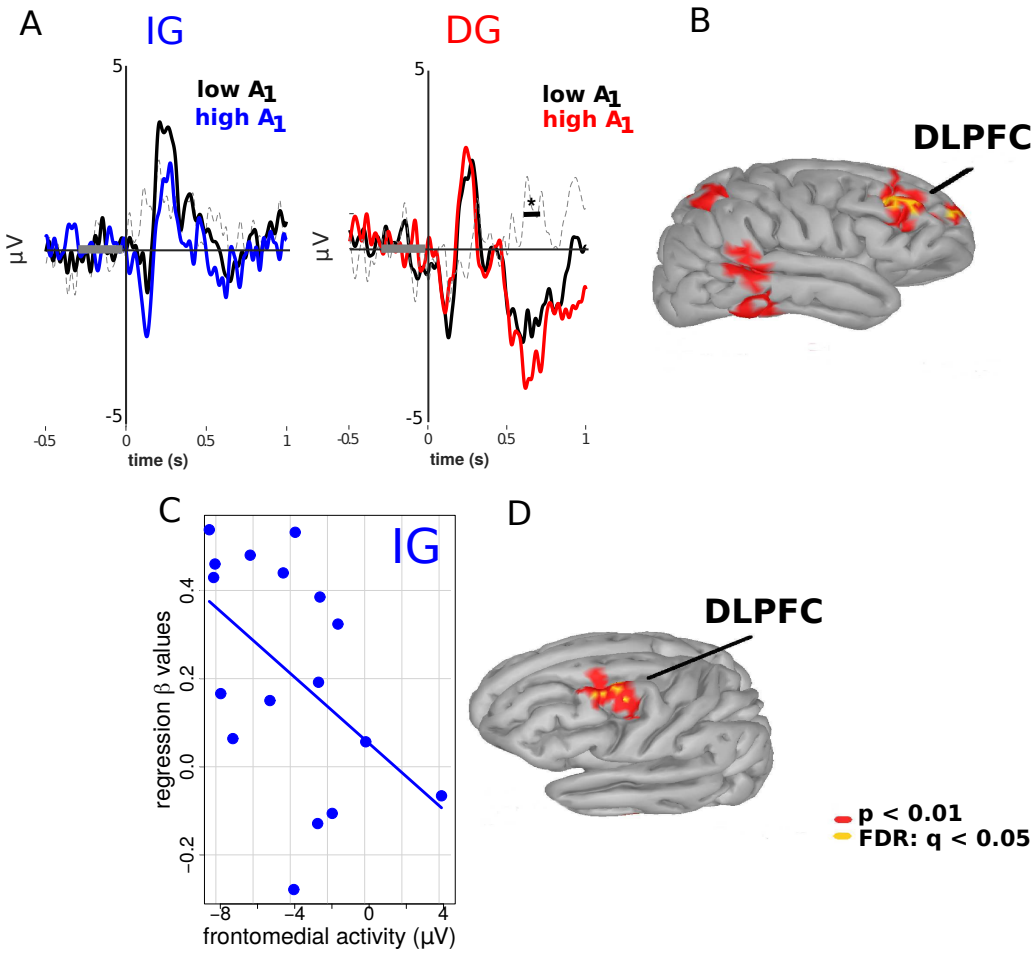


Figure 11: FMN encodes the magnitude of A_1 under the DG condition. **A.** Left: ERPs of subjects when they received an allocation in the IG. Blue and black lines represent subjects' ERP when values of A_1 were above and below €6, respectively. Right: ERPs of subjects for the DG condition. Red and black lines represent subjects' ERP when values of A_1 were above and below €6, respectively. Significant differences were obtained in a time windows between 550 and 650 ms after stimulus onset, specifically for this condition. **B.** Cortical source projections (FDR: $q < 0.05$; $p < 0.01$, uncorrected). * indicates p < 0.05. Subjects' average FMN predicts their responses to trust. **C.** Regression of average ERPs in the frontomedial cluster (X-axis) versus the β-values obtained from Equation 1 (Y-axis) for the IG (intentions-based) condition ($\beta = -0.037$; $p = 0.045$). **D.** Cortical source projections (FDR: $q < 0.01$; $p < 0.005$, uncorrected).

et al., 2003), for high amount of received trust, (ii) an area where altruism and reciprocity were indistinguishable, and (iii) an area of negative reciprocity, for low

levels of trust. Thus, we found that reciprocity is positive only for high amounts of trust, but turns negative (less than expected for altruism) for lower amounts. Positive reciprocity has been discussed elsewhere (Cox, 2004; Ashraf et al., 2006; McCabe et al., 2003). Negative reciprocity could be interpreted as indicating that subjects feel social betrayal if the amount of trust received is less than what they expect from a certain social norm of expected trust (Coleman, 2000; Gouldner, 1960) – similar to the results obtained in the UG (Güth et al., 1982; Kaltwasser et al., 2016; Sanfey et al., 2003; Billeke et al., 2013; 2014; even in the third-party version of the game; Civai et al., 2010).

At the neurobiological level, a late frontal negativity was more negative when subjects were notified about the amount they had to share in the outcome- versus the intention-based condition. Cortical source estimations for the ERP differences projected to ACC, DLPFC, and TPJ, structures participating in mentalisation and cognitive control networks. ACC participates in cognitive control processes in both social (Apps et al., 2016; Apps and Ramanini, 2014) and non-social (Kolling et al., 2016; Hauser et al., 2014; Shenhav et al., 2013; 2014; Ullsperguer et al., 2014; Holroyd and Coles, 2002) scenarios. ACC has also been associated with the maintenance of trust and reciprocal interactions (King-Casas et al., 2005; Baumgartner et al., 2008) and its activity is modulated by the "prosocial" neuropeptide oxytocin (Baumgartner et al., 2008; Aspé-Sánchez et al., 2016). In addition, theta activity projecting to ACC might reflect a behavioural heuristic adaptation to the behavior of others (Billeke et al., 2014). TPJ has been shown to be important in the control of selfish impulses (Hutcherson et al., 2015) and the valuation of others' outcomes (Hutcherson et al., 2015; Hein et al., 2016). During social interactions, TPJ alpha activity correlated with the anticipation of the other's behaviour, and with the use

of mentalisation in planning future actions (Hill et al., 2017; Melloni et al., 2016, Soto-Icaza et al., 2019). DLPFC, on the other hand, is involved in strategic and normative decision making (Baumgartner et al., 2011; Chang et al., 2011; Yamagishi et al., 2016), and when matching the other player's expectations about a social interaction (Chang et al., 2011).

Specifically, we found that under the outcome-based trials this late frontal negativity was more negative when subjects received greater allocations, with scalp activity projecting to DLPFC. In this respect, two studies have found that a dictator's unfair offer elicits, in the recipients, a more negative ERP (specifically a feedback-related negativity) than does a fair offer (Wu et al., 2011) – even in the third-person version of the game (i.e., when subjects observe others receiving the allocation; Mothes et al., 2016). In contrast, here subjects processed the allocations received in the outcomes-based condition not as recipients, but as dictators, with greater allocations from the COM implying a greater endowment to choose to share with their human partner. Since greater frontal negativity was observed when subjects received high versus low amounts, it is possible that subjects recruited more cognitive control in order to inhibit the impulse to be selfish – similar to situations in which feedback-related negativity amplitude indicates more cognitive control (Ullsperguer et al., 2014; Holroyd and Coles, 2002). Consistent with this interpretation, the associated activity projected to structures involved in the normative network (specifically, DLPFC), arguably in order to override subjects' temptations to keep a greater amount for themselves (Baumgartner et al., 2011; Chang et al., 2011; Yamagishi et al., 2016).

In the IG conditions, by contrast, we did not observe an association between the amount received (now from a partner believed to be human) and the amplitude of the late frontal negativity. This could be explained because reciprocal behaviour is

more common in nature (Franzen and Pointner, 2013; Johnson and Mislin, 2011): we trust and repay trust from conspecifics everyday, thus requiring less cognitive control. An additional – and complementary – interpretation could be that allocations in the intentions-based conditions are processed in a more heuristic fashion. Much research has indicated that, in social dilemmas, subjects apply different rules (heuristics) without necessarily recruiting neurophysiological markers of cognitive control (Billeke et al., 2014; Hill et al., 2017, Yamagishi et al., 2016). Interestingly, these different rules correlated with activity in medial prefrontal regions, DLPFC and TPJ (Billeke et al., 2014; Hill et al., 2017, Melloni et al., 2016, San Martín et al., 2016, Yamagishi et al., 2016). The use of a heuristic strategy in our experiments is supported by the fact that subject's average ERP amplitudes showed a significant negative correlation with how much they increased (decreased) their reciprocity when facing more (less) trust, with this correlation projecting to the DLPFC, which could indicate the use of normative rules (Lim et al., 2016). Neurophysiologically, negative reciprocity has previously been investigated only in association with the rejection of unfair offers in the UG, in which brain areas related to cognitive control and normative decision making (such as the DLPFC) correlate with the rejection of unfair offers (Sanfey et al., 2003). Our results suggest that DLPFC activity is associated with the observed behaviour in complex social interactions, such as those requiring reciprocity.

In summary, the use of the hybrid DIG revealed that reciprocal behaviour is positive for high amount of received trust, but negative for low amounts. In addition, altruism and reciprocity evoke different activity in brain networks involved in mentalising and cognitive control, which are involved in the inhibition of selfish behaviour, and the processing of the internal states, the perspectives, and even the monitoring of the performance of others, such as during (vicarious) reward prediction errors (Apps et

al., 2016; Wittman et al., 2017; Amodio and Frith, 2006). Thus, our findings expand our current knowledge about the relationship between brain networks involved in mentalising and cognitive control processes, social preferences, and complex social behaviours.

Acknowledgments

This work was supported by the Chilean National Scientific and Technological Research Commission (CONICYT) FONDECYT (1181295 to PB, and 1180403 to JE); and the Centro Interdisciplinario de Neurociencia de Valparaiso (CINV) Millennium Institute (P09-022-F to JE), supported by the Millennium Scientific Initiative of the Ministerio de Economia, Fomento y Turismo. MA thanks to Ricardo Guzmán, Ph.D. (Centro de Investigación en Complejidad Social) for his comments and suggestions on the experimental design and regression models. Correspondence should be addressed to: mauricio.aspe.s@gmail.com

Conflicts of interest

The authors declare no conflicts of interest.

Code and data availability statement

All data and script used for data analyses are free available on http://neuroCICS.udd.cl

4.6 References

Abu-Akel, A., & Shamay-Tsoory, S. (2011) Neuroanatomical and neurochemical bases of theory of mind. Neuropsychologia, 49(11):2971–2984.

Amodio, D. M., & Frith, C. D. (2006) Meeting of minds: the medial frontal cortex and social cognition. Nature Reviews Neuroscience, 7(4):268–77.

Andreoni, J., & Miller, J. (2003) Giving according to GARP: An experimental test of the consistency of preferences for altruism. Econometrica, 70(2):737–753.

Apps, M. A. J., & Ramnani, N. (2014) The Anterior Cingulate Gyrus Signals the Net Value of Others' Rewards. Journal of Neuroscience, 34(18):6190–6200.

Apps, M. A. J., Rushworth, M. F. S., & Chang, S. W. C. (2016) The Anterior Cingulate Gyrus and Social Cognition: Tracking the Motivation of Others. Neuron, 90(4):692–707.

Ashraf, N., Bohnet, I., & Piankov, N. (2006) Decomposing trust and trustworthiness. Experimental Economics, 9(3):193–208.

Aspé-Sánchez, M., Moreno, M., Rivera, M. I., Rossi, A., & Ewer, J. (2015) Oxytocin and Vasopressin Receptor Gene Polymorphisms: Role in Social and Psychiatric Traits. Frontiers in Neuroscience, 9(January):510.

Axelrod, R. & Hamilton, W. D. (1981) The evolution of cooperation. Science 211, 1390—1396.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008) Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron, 58(4):639–50.

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., & Fehr, E. (2011) Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. Nature Neuroscience, 14(11), 1468–1474.

Berg, J., Dickhaut, J., & McCabe, K. (1995) Trust, reciprocity, and social history. Games and Economics Behaviour, (10):122–142.

Billeke, P., Zamorano, F., Cosmelli, D., & Aboitiz, F. (2013) Oscillatory Brain Activity Correlates with Risk Perception and Predicts Social Decisions. Cerebral Cortex, 23(12), 2872–2883.

Billeke, P., Zamorano, F., Chavez, M., Cosmelli, D., & Aboitiz, F. (2014) Functional cortical network in alpha band correlates with social bargaining. PLoS ONE, 9(10).

Billeke, P., Soto-Icaza, P., Aspé-Sánchez, M., Villarroel, V., & Rodríguez-Sickert, C. (2017) Valuing others: evidence from economics, developmental psychology, and neurobiology. In Neuroscience and Social Science (pp. 21-45). Springer, Cham.

Bolton, G. & Ockenfels, A. (2000) ERC: A theory of equity, reciprocity, and competition. American Economic Review, 90(1):166–193.

Camerer, C. F. (2003) Behavioural Game Theory. New York. Princeton University

Press.

Cavanagh, J. F., & Frank, M. J. (2014) Frontal theta as a mechanism for cognitive control. Trends in Cognitive Sciences, 18(8), 414–421.

Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011) Triangulating the neural, psychological, and economic bases of guilt aversion. Neuron, 70(3):560–72.

Civai, C., Corradi-Dell'Acqua, C., Gamer, M., & Rumiati, R. I. (2010) Are irrational reactions to unfairness truly emotionally-driven? Dissociated behavioural and emotional responses in the Ultimatum Game task. Cognition, 114(1):89–95.

Clerc, M., Gramfort A., Olivi, E., & Papadopoulo, T. (2010) The symmetric BEM: bringing in more variables for better accuracy. In: Supek S, Sušac A, editors. 17th International Conference on Biomagnetism Advances in Biomagnetism – Biomag 2010. Dubrovnik (Croatia).

Coleman, J. (2000) Foundations of Social Theory. Cambridge (UK). Harvard University Press.

Costa, A., Torriero, S., Oliveri, M., & Caltagirone, C. (2008) Prefrontal and temporoparietal involvement in taking others' perspective: TMS evidence. Behavioural Neurology, 19(1-2):71–74.

Cox, J. C. (2004) How to identify trust and reciprocity. Games and Economic Behavior, 46(2):260–281.

Decety, J. & Lamm, C. (2007) The Role of the Right Temporoparietal Junction in Social Interaction: How Low-Level Computational Processes Contribute to Meta-Cognition. The Neuroscientist, 13(6):580–593.

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(11):1611–8.

Delorme, A., & Makeig, S. (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 134(1), 9-21.

Fehr, E., & Schmidt, K. (2001) Theories of fairness and reciprocity - Evidence and economic applications. CESifo Working Paper Series, (1214).

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994) Fairness in simple bargaining experiments. Games Econ. Behav., (6):347–369.

Franzen, A., & Pointner, S. (2013) The external validity of giving in the dictator

game. Experimental Economics, 16(2), 155-169.

Frith, C. D. & Frith, U. (1999) Interacting Minds - A Biological Basis. Science, 286(5445):1692–695.

Gouldner, A. W. (1960) The norm of reciprocity: A preliminary statement. American sociological review, 161-178.

Güth, W., Schmittberger, R., & Schwarze, B. (1982) An experimental analysis of ultimatum bargaining. Journal of Economic Behavior & Organization, 3(4), 367-388.

Hamilton, W. D. (1996) Narrow Roads of Gene Land Vol. 1. New York. Freeman.

Hauser, T. U., Iannaccone, R., Stämpfli, P., Drechsler, R., Brandeis, D., Walitza, S., & Brem, S. (2014) The feedback-related negativity (FRN) revisited: New insights into the localization, meaning and network organization. NeuroImage, 84, 159–168.

Hein, G., Morishima, Y., Leiberg, S., Sul, S., & Fehr, E. (2016) The brain's functional network architecture reveals human motives. Science, 351(6277):1074–1078.

Hill, C. A., Suzuki, S., Polania, R., Moisa, M., O'Doherty, J. P., & Ruff, C. C. (2017) A causal account of the brain network computations underlying strategic social behavior. Nature Neuroscience, 20(8).

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(4):679–709.

Hutcherson, C. A., Bushong, B., & Rangel, A. (2015) A Neurocomputational Model of Altruistic Choice and Its Implications. Neuron, 87(2):451–463.

Johnson, N. D. & Mislin, A. (2011) Trust games: A meta-analysis. Journal of Economic Psychology, 32(5):865–889.

Kaltwasser, L., Hildebrandt, A., Wilhelm, O., & Sommer, W. (2016) Behavioral and neuronal determinants of negative reciprocity in the ultimatum game. Social Cognitive and Affective Neuroscience, (April):1–10.

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.

Kolling, N., Wittmann, M. K., Behrens, T. E. J., Boorman, E. D., Mars, R. B., & Rushworth, M. F. S. (2016) Value, search, persistence and model updating in anterior cingulate cortex. Nature Neuroscience, 19(10):1280–1285.

Larrain-Valenzuela, J., Zamorano, F., Soto-Icaza, P., Carrasco, X., Herrera, C., Daiber, F., Aboitiz, F., & Billeke, P. (2017) "Theta and Alpha Oscillation Impairments in Autistic Spectrum Disorder Reflect Working Memory Deficit." Scientific Reports 7 (1).

Lim, S.-L., Cherry, J. B. C., Davis, A. M., Balakrishnan, S. N., Ha, O.-R., Bruce, J. M., & Bruce, A. S. (2016) The child brain computes and utilizes internalized maternal choices. Nature Communications, 7(1), 11700.

Lin F-H, Witzel T, Ahlfors SP, Stufflebeam SM, Belliveau JW, & Hämäläinen MS. (2006) Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. Neuroimage. 31:160–171.

Maris, E. & Oostenveld, R. (2007) Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods, 164(1):177–190.

McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003) Positive reciprocity and intentions in trust games. Journal of Economic Behavior and Organization, 52(2):267–275.

Melloni, M., Billeke, P., Baez, S., Hesse, E., De la Fuente, L., Forno, G., & Slachevsky, A. (2016) Your perspective and my benefit: multiple lesion models of self-other integration strategies during social bargaining. Brain, 139(11), 1-19.

Mothes, H., Enge, S., & Strobel, A. (2016) The interplay between feedback-related negativity and individual differences in altruistic punishment: An EEG study. Cognitive, Affective, & Behavioral Neuroscience, 16:276–288.

Nowak, M. A. & Sigmund, K. (2005) Evolution of indirect reciprocity. Nature, 437(October):1291–1298.

Peirce, Jonathan W. (2009) PsychoPy - Psychophysics software in Python. Journal of Neuroscience Methods. 162 (1-2): 8–13.

R Core Team. (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

San Martín, R., Kwak, Y., Pearson, J. M., Woldorff, M. G., & Huettel, S. A. (2016) Altruistic traits are predicted by neural responses to monetary outcomes for self vs charity. Social Cognitive and Affective Neuroscience, 11(6), 863–876.

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(5626):1755–8.

Saxe, R. R., Whitfield-Gabrieli, S., Scholz, J., & Pelphrey, K. A. (2009) Brain re-

gions for perceiving and reasoning about other people in school-aged children. Child Development, 80(4):1197–1209.

Selten, R. (1965) Spieltheoretische behandlung eines oligopolmodells mit nachfragetr agheit: Teil I: Bestimmung des dynamischen preisgleichgewichts. Zeitschrift für die gesamte Staatswissenschaft, 121(2):301–324.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013) The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron, 79(2):217–40.

Shenhav, A., Cohen, J. D., & Botvinick, M. (2016) Dorsal anterior cingulate cortex and the value of control. Nature Neurosci, 19(10):1286–1291.

Soto-Icaza, P., Vargas, L., Aboitiz, F., & Billeke, P. (2019) Beta oscillations precede joint attention and correlate with mentalization in typical development and autism. Cortex, 113, 210-228.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011) Brainstorm: A User-Friendly Application for MEG/EEG Analysis. Computational Intelligence and Neuroscience, Vol. 2011, Article ID 879716.

Trivers, R. (1985) Social evolution (No. 156 T7).

Ullsperger, M., Fischer, A. G., Nigbur, R., & Endrass, T. (2014) Neural mechanisms and temporal dynamics of performance monitoring. Trends in Cognitive Sciences, 18(5):259–267.

van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A. R. B., & Crone, E. A. (2009) What motivates repayment? Neural correlates of reciprocity in the Trust Game. Social cognitive and affective neuroscience, 4(3):294–304.

Wilson, E. O. (1975) Sociobiology. Cambridge. Harvard Univ. Press.

Wittmann, M., Kolling, N., Faber, N., Scholl, J., Nelissen, N., & Rushworth, M. (2016) Self-Other Mergence in the Frontal Cortex during Cooperation and Competition. Neuron, 91(2):482–493.

Wu, Y., Leliveld, M. C., & Zhou, X. (2011) Social distance modulates recipient's fairness consideration in the dictator game: an ERP study. Biological Psychology, 88(2-3):253–62.

Yamagishi, T., Takagishi, H., Fermin, A. de S. R., Kanai, R., Li, Y., & Matsumoto, Y. (2016) Cortical thickness of the dorsolateral prefrontal cortex predicts strategic choices in economic games. Proceedings of the National Academy of Sciences, 113(20), 5582–5587.

5 The neuroeconomics of trustworthiness during childhood, adolescence and early adulthood

Ongoing article, written by:

Mauricio Aspé-Sánchez[1], Nicolás Ferrer[1], Josefina Larraín[1], Pablo Billeke[1], Francisco Zamorano[1] & Carlos Rodríguez-Sickert[2]

[1] División de Neurociencias (NeuroCICS), Centro de Investigación en Complejidad Social, Facultad de Gobierno, Universidad del Desarrollo, Santiago, Chile.

[1] Centro de Investigación en Complejidad Social (CICS), Facultad de Gobierno, Universidad del Desarrollo, Santiago, Chile.

5.1 Introduction

In my previous work, I found that subjects increased their prosocial behaviour depending on the amount of trust they received. Interestingly, a subject's late frontal negativity (electroencephalographic component that accounts for cognitive control and performance monitoring) differed between conditions, predicting responses to trust specifically in intentions-based trials. In addition, I found that the temporoparietal junction and the dorsomedial and dorsolateral prefrontal cortices – structures related with mentalising and cognitive control – are the cortical sources of this activity (Aspé-Sánchez et al., submitted to *Frontiers in Social Psychology*; see also Section 5).

In this line, research has proposed that social abilities could be the product of the joint function of 3 neurobiological systems (Declerck et al., 2013). The first is the reward system, mainly conformed by dopaminergic projections coming from the ventral tegmentum, in the midbrain, to areas related with motivation and the processing of incentives (Cummings, 1995; Tekin and Cummings, 2002). The second is the cognitive control system, which allows strategic decision-making when incentives are contradicting (Rilling, King-Casas & Sanfey, 2008). Finally, the social cognition system allows us to identify agents of biological relevance (for instance, conspecifics) and infer their future actions (Van Overwalle, 2011). Those three system are not clearly developed in the new-born, maturating during lifetime (Atzil et al., 2018). For instance, the social cognition system (related to the processes of the mentalising) is not completely developed until the cortical frontal component is fully mature, which would occur during late adolescence (Sowell et al., 2003; Giedd, 2004; Giedd et al., 1999; Sowell et al., 1999; Gogtay et al., 2004). The same runs for the cognitive control system (Westlye et al., 2010), whose development culminates in advanced

stages of adulthood (Rubia et al., 2006). This cortical region plays a fundamental role in the individual's social adaptation, for it participates directly in controlling inhibitory processes, strategic planning, decision-making, and the adaptation of social behavior to different circumstances (Yang and Raine, 2009).

Prosocial behaviours also differ between different stages of an individual's development. Belli et al. (2012) show that adults generally reciprocated their partners' trust even when little trust was deposited in them, or when the trust relationship was simply broken. This is not true for adolescents, who were mainly outcome-based motivated (Belli et al., 2012). Developmental psychologists tend to understand the role age plays in social dilemmas by emphasizing that the internalization of social rules is a gradual process (Kolhberg, 1971; Hy and Loevinger, 1996). For instance, behavioral inhibition – modulated by the neuroanatomical development of the cognitive control systems – plays a crucial role in the implementation of fair behavior in bargaining games (Steinbis et al., 2013). In addition, a study conducted with children between ages 6 and 13 demonstrated that developmental changes in the left DLPFC cortex were significantly correlated with the control of impulses and capacity of negotiation, which leads to the idea that young children's rejection of unfair offers would be determined by a poor behavioral control, rather than by a lack of comprehension of norms or social abilities (Steinbeis et al., 2012). These results suggest that the bases of selfish or antisocial behavior in children do not arise from a lack of notions about what is correct and what is not, but rather from the inability to inhibit behaviours centered in personal benefit when faced to situations that present strong, immediate incentives for pro-self choices.

In this work, we hypothesize that the degree of development of each of the neurocognitive systems above mentioned modulate the weight attached to that system in

social decison-making processes. In order to test this hypothesis, we comparatively studied the cognitive and behavioral response of subjects when trust is placed in them, using a Go-Nogo and a Dictator/Investment game (Cox, 2005, Aspé-Sánchez et al., submitted to *Frontiers in Social Psychology*), respectively, in three different samples: children (10-12), adolescents (14-16) and young adults (20-22). Bayesian analysis using a neurocognitive model shows that adults display greater inhibitory cognitive control and that this modulates differently the subjects' prosocial behaviour across different ages. Our results represent an important contribution to understand the evolution of social behavior, with implications ranging from educational policies to investigation methodologies.

5.2 Methods

5.2.1 Participants

We studied three different age ranges, using two behavioural protocols: a task of behavioral inhibition (Go-Nogo) and a task of prosocial preferences (DIG, see below). Participants were divided in three groups: the first corresponded to children between 10-12 years (children group), the second to adolescents aged between 14-16 years (adolescent group) and the third to adults, aged 20 to 22 years (adult group). Participants was evaluated by a psychologist (WISC-R, mini-KID and WAIS, SCAN) in order to discard confounding comorbidities and ensure an IQ homogeneous sample. All subjects and their parents (underage) signed an informed consent, approved by the Ethics Committee of the University, to voluntarily participate in this study.

5.2.2 Experimental protocol.

Go-Nogo. Participants performed a simple version of the Go-Nogo task. The inter-stimuli interval (ISI) was drawn from uniform distributions in [500 - 800] ms. Stimulus duration was of 300 ms, indistinctly for Go and Nogo trials. Each participant performed two blocks of 300 trials. This paradigm consists of a serial presentation of screen centered green or red colored circles (3.5 degrees of visual arc) for Go and Nogo stimuli, respectively. Nogo trials was preceded by sequences of 1, 3, 5, 7 Go trials, in order to exacerbate the response automation. By this way, more inhibition processing was required after a longer Go stimuli sequence to refrain from pressing the button. Stimuli was presented in a LCD monitor positioned 57 cm in front of the subject. The subjects was instructed to press a button as fast as possible after a Go stimulus (green circle) and to restrain after a Nogo stimulus (red circle). Two consecutive Nogo stimuli was never presented. The tasks was programmed in

Presentation 13.0 (Neurobehavioral Systems, Inc.). We measured the reaction times for responses to Go and Nogo stimuli, and the number of correct answers for each participant, in order to evaluate the subject's behavioral inhibition performance.

Dictator/Investment game. In our modified version of Berg et al (1995)'s Investment game (see Aspé-Sánchez et al., submitted to Frontiers in Social Psychology, and Section 5), the trustees face allocation A_1 in the range $[0, 10]$, comming from another human (HUM trials) or a computer (COM trials). In order to differentiate between intention-based and outcome-based motivations of the trustee, two alternative treatments are considered. In the baseline treatment (HUM), it is the truster who decides the level of investment. In the alternative treatment (COM), the decision is randomly taken by a computer (the investment game becomes a set of dictator game in terms of the payoff structure).

At the end of each session experimental subjects was given financial compensation proportional to the amount of money collected during the game.

5.2.3 Model

Inhibitory cognitive control. I defined a hierarchical model were correct ommissions ($co_{(t)}$) is the result of a Bernoulli process $co \sim dBern(\theta)$ with parameter $\theta \sim dBeta(\alpha, \beta)$:

$$\alpha, \beta \sim U(0, 1) \tag{3}$$

$$co \sim dBern(\theta) \tag{4}$$

$$\theta \sim dbeta(\alpha, \beta) \tag{5}$$

Thus, θ represents a distribution of each subjects inhibitory cognitive control.

Prosocial orientation. I further define a hierarchical robust lineal regression model to account for individuals prosocial orientation, with:

$$\mu_{(s)} \sim N(0, \epsilon) \tag{6}$$

$$\sigma \sim U(\epsilon, 10) \tag{7}$$

$$\nu \sim Exp(1/29) + 1 \tag{8}$$

$$\epsilon = 1 \times 10^{-5} \tag{9}$$

$$\beta_{0(s,t)} \sim t(\mu_{(s)}, 1/\sigma^2, \nu) \tag{10}$$

$$pt_{(s,t)} \sim Bern(\psi_{(s,t)}) \tag{11}$$

with:

$$\psi_{(s,t)} = \frac{1}{1 + \exp^{-\rho}} \tag{12}$$

$$\rho_{(t)} = \beta_{0(s)} + \beta_{1(s)} H_{(t)} + \beta_{2(s)} \hat{A}_{1(t)} + \beta_{3(s)} \hat{A}_{1(t)} H_{(t)} + \beta_{4(s)} \theta_{(s)} + \beta_{5(s)} \theta_{(s)} H_{(t)} \tag{13}$$

where $H_{(t)}$ is a dummy variable equal to 1 if the subject is facing an human allocation or a COM allocation, $\hat{A}_{1(t)}$ is the normalised allocation (A_1) received ($A1/10$), $\theta_{(s)}$ is the subjects' cognitive control capability given by the Bayesian model of Equation

6-13, and $pt = 1$ if $A_2 \geq A_1$, and 0 otherwise.

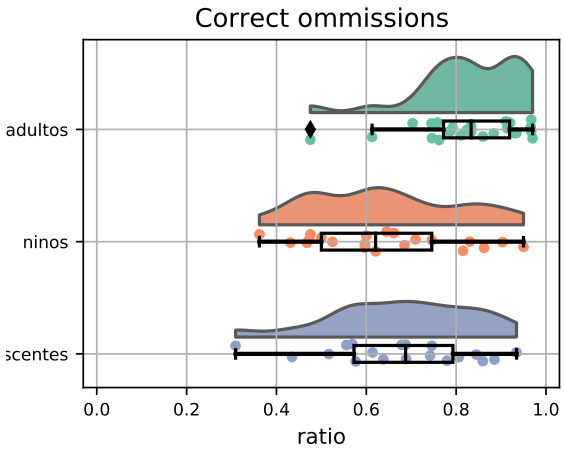
5.3 Results

5.3.1 Cognitive control

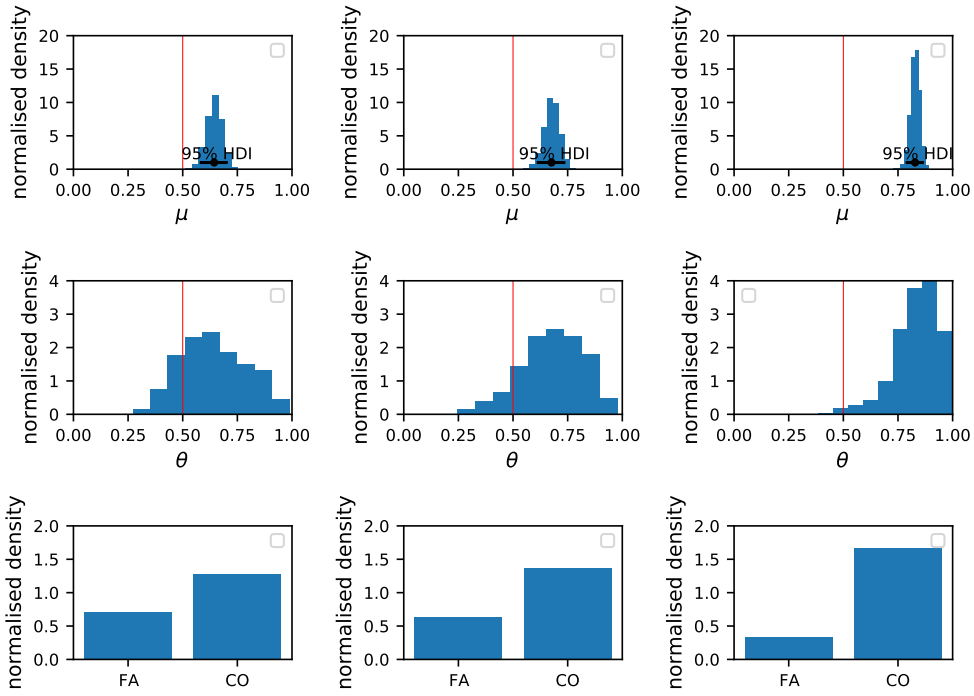
I first quantified participants inhibitory cognitive control capability θ as the ratio between subjects' correct ommissions and total nuber of non-target trials (Figure 12). Observed values suggest that adults present greater θ than children and adolescents (children: $n = 21$; $\mu = 0.64$; $\sigma = 0.17$; adolescent: $n = 19$; $\mu = 0.68$; $\sigma = 0.16$; adults: $n = 27$; $\mu = 0.83$; $\sigma = 0.12$). This is confirmed by the hierarchical Bayesian model specified in Methods, where adults present an average θ for the Markov chains significantly different from 0.5 ($\mu = 0.83$; $\sigma = 0.12$; 95% HDI $= [0.59, 0.98]$), while this is not the case for children ($\mu = 0.64$; $\sigma = 0.17$; 95% HDI $= [0.38, 0.92]$) and adolescent ($\mu = 0.67$; $\sigma = 0.16$; 95% HDI $= [0.37, 0.91]$; Figure 12). This results suggest that only adults perform better than expected by chance.

5.3.2 Prosocial orientation

I next modeled subjects' prosocial orientation as the parameter ψ of the Bernoulli process $pt \sim B(\psi)$, where $pt = True$ if $\hat{A}_2 \geq T$, with $T = 1/3$ (*i.e.*, trials where subjects sent at least A_1). Thus, ψ represents the point in the underlying neurocognitive process that yields in an outcome where $A_2 \geq A_1$. Figure 2 shows the subjects' mean ψ-values for the Markov chains (black points) estimated with the Bayesian model, plus 100 data points inside the 95% HDI of the same subject $\psi \sim dbeta$ distribution (blue points). This is depicted for both the COM and the HUM conditions. Qualitative analysis of the data suggests that, while children and adolescents yield mainly along the $x = y$ diagonal, adults *(i)* decrease their proportions of means – but a



(a) Correct ommissions. Subjects average correct ommissions ratio, defined as # correct ommissions / # non-target trials. The plot displays, for each condition, the smoothed distribution (colored areas) of individual means (points), plus a boxplot.



(b) Markov chains distributions for θ. (Top) Histogram for 1000 samples from 5 chains after 1000 adaptation steps, for subjects' θ parameter (blue bars) and a prior that is not updated with the data (black lines). The 95% HDI is shown at the bottom of each plot. Adults' θ distribution differs from 0.5. (Bottom) Histogram of false alarms (FA) and correct ommissions (CO) from subjects' observed data. Black lines show the distribution estimated from the posterior $CO \sim Bern(\theta)$.

Figure 12: Mean inhibitory cognitive control parameter θ across the different groups.

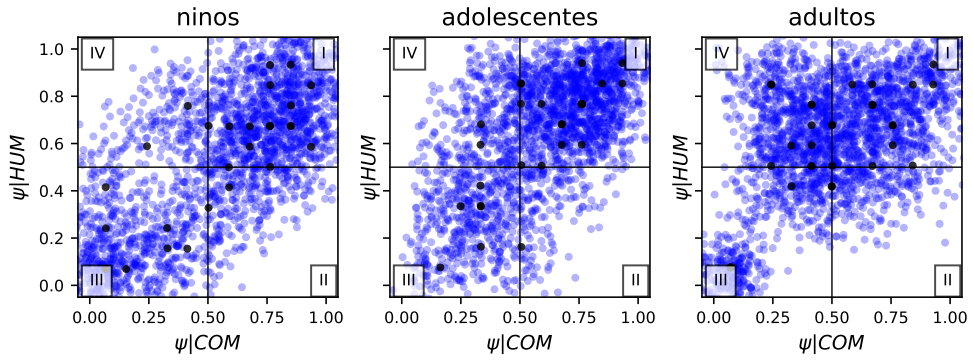


Figure 13: Prosocial orientation. Subjects' mean ψ-values for the Markov chains estimated with the Bayesian model, plus 100 data points inside the 95% HDI of the same subject $\psi \sim$ *dbeta* distribution (blue points). Quadrants allows to clasify them among different types of players: I: Prosocials; II: Altrusitic prosocials; III: Means; IV: Reciprocal prosocials.

cluster of means emerges (Quadrant III, Figure 2), and *(ii)* increase their proportion of reciprocal prosocials (Quadrant IV, Figure 13).

In order to investigate the role of inhibitory cognitive control on prosocial behaviour, I model the prosocial trials $pt \in [0, 1]$ as a Bernoulli arising from the subject ψ logistic regression:

$$\psi_{(s,t)} = \frac{1}{1 + \exp^{-\rho}}$$

$$\rho_{(t)} = \beta_{0(s)} + \beta_{1(s)} H_{(t)} + \beta_{2(s)} \hat{A}_{1(t)} + \beta_{3(s)} \hat{A}_{1(t)} H_{(t)} + \beta_{4(s)} \theta_{(s)} + \beta_{5(s)} \theta_{(s)} H_{(t)}$$

where $H_{(t)}$ is 1 if subjects received A_1 from another human, and 0 if where received from COM; $\theta_{(s)}$ is subject inhibitory cognitive control parameter, as obtained from the model.

The result (Figure 14) show that, effectively, the amount received and the cognitive

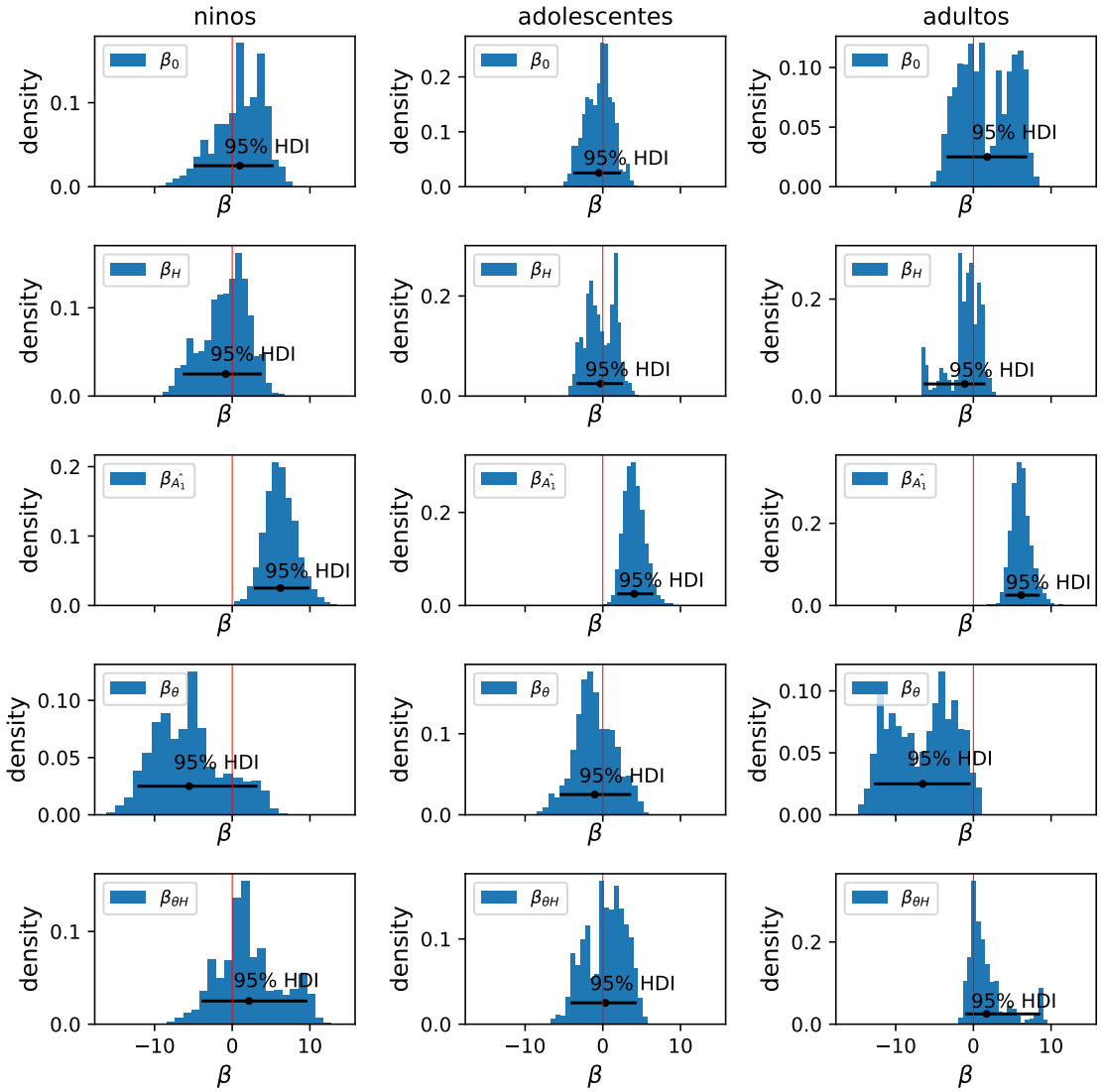


Figure 14: Markov chains distributions for significants β. Histogram correspond to 1000 samples from 5 chains after 1000 adaptation steps, according with the model presented in Equation 6-13. The 95% HDI is shown at the bottom of each plot.

control parameter influences prosocial behaviour in a different fashion among groups. Specifically, while all groups increase in their prosocial orientation as their receive a greater A_1 (children: $\beta_{\hat{A_1}}$: $\mu = 4.90$; $\sigma = 1.18$; 95% HDI $= [2.99, 6.80]$; adolescents: $\beta_{\hat{A_1}}$: $\mu = 3.03$; $\sigma = 0.78$; 95% HDI $= [1.79, 4.34]$; adults: $\beta_{\hat{A_1}}$: $\mu = 4.99$; $\sigma = 0.68$; 95% HDI $= [3.84, 6.15]$), cognitive control θ display a role only in the children and adults group (children: β_θ: $\mu = -7.55$; $\sigma = 4.16$; 95% HDI $= [-15.09, -1.75]$; adults: β_θ: $\mu = -6.51$; $\sigma = 3.01$; 95% HDI $= [-11.17, -1.58]$), without displaying an influence in adolescents (β_θ: $\mu = 0.08$; $\sigma = 2.77$; 95% HDI $= [-5.04, 4.16]$). I failed to found significant differences in ψ when subject face a HUM versus a COM counterpart (β_H: $\mu = -0.73$; $\sigma = 2.98$; 95% HDI $= [-4.13, 5.46]$ for adults, for instance) or the interaction between θ and HUM trials ($\beta_{\theta H}$: $\mu = 1.38$; $\sigma = 3.55$; 95% HDI $= [-4.58, 7.43]$ for children, for instance). It is noteworthy, however, that adults show a marked bimodal distribution for the interaction of θ and HUM trials, suggesting that there could be two underlying populations (Figure 14; see also the "means-cluster" of Figure 13).

Finally, the results in Figure 15 show the distribution of the ψ and pt values given the β parameters specified below.

5.4 Discussion

Currently, it is an open question whether cognitive control is a factor influencing prosocial behaviour, and in which direction it impact: Is prosocial behaviour a heuristic, more related with an instinctive behaviour? Or the cognitive control system is required for prosocial manifestations? The former implies that greater cognitive control is associated with less prosocial behaviour, adding an inhibitory component to the prosocial heuristic, while the latter implies that less cognitive control is as-

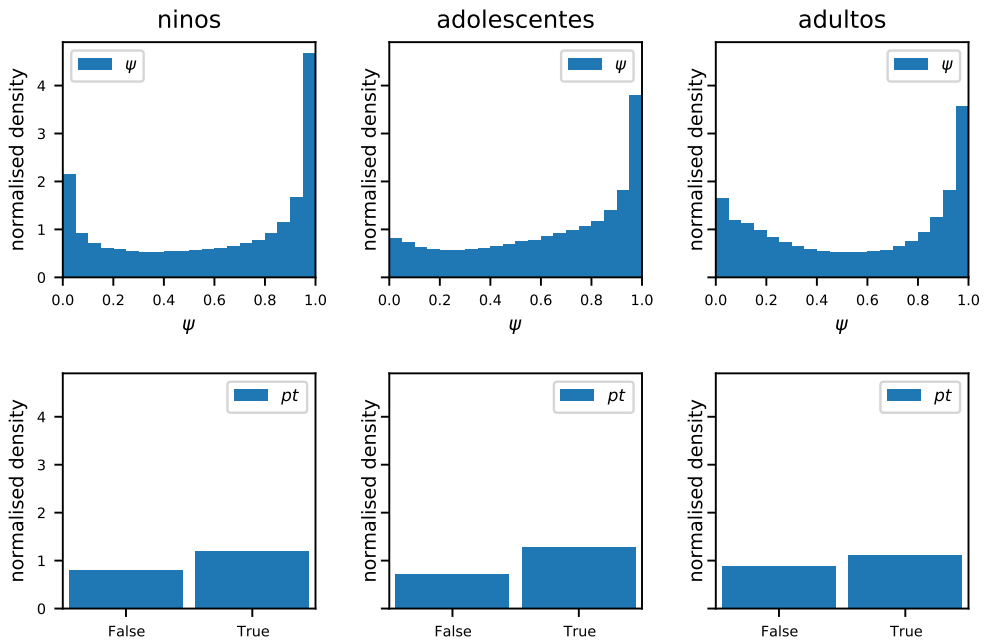


Figure 15: Markov chains distributions for ψ and prosocial trials (pt). (Top) Histogram for 1000 samples from 5 chains after 1000 adaptation steps, for subjects' ψ parameter (blue bars) and a prior that is not updated with the data (black lines). (Bottom) Histogram of prosocial trials pt. Black lines show the distribution estimated from the posterior $pt \sim Bern(\psi)$.

sociated with greater prosocial behaviour, given that the heuristic is to be pro-self and the cognitive inhibitory component acts increasing prosocial allocations by, for instance, fit the behaviour to social norms (Yamagishi et al., 2016).

In this work I show that adults *(i)* present a greater inhibitory cognitive control, *(ii)* that subjects' PO increases when they received a greater allocation, and that *(iii)* in children and adults (but not in adolescents) greater cognitive control capabilities also implies lower POs. This shed evidence supporting that cognitive control acts reducing prosocial behaviour and that the heuristic, 'control-free" behaviour is prosocial, but introduce complexity on the development of prosocuiality and its interplay with the developmental course of the cognitive control system (see Introduction). We

hope to conduct further investigations to solve these open questions.

5.5 Bibliography

Bartzokis G, Beckson M, Lu PH, Nuechterlein KH, Edwards N, Mintz J (2001) Age-related changes in frontal and temporal lobe volumes in men: a magnetic resonance imaging study. Archives of General Psychiatry 58:461-465.

Baumgartner T, Heinrichs M, Vonlanthen A, Fischbacher U, Fehr E (2008) Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron 58:639-650.

Bechara A, Damasio H, Tranel D, Damasio AR (2005) The Iowa Gambling Task and the somatic marker hypothesis: some questions and answers. Trends Cogn Sci 9:159-162.

Belli SR, Rogers RD, Lau JY (2012) Adult and adolescent social reciprocity: Experimental data from the Trust Game. J Adolesc.

Camerer CF (2003) Psychology and economics. Strategizing in the brain. Science 300:1673-1675.

Casey BJ, Castellanos FX, Giedd JN, Marsh WL, Hamburger SD, Schubert AB, Vauss YC, Vaituzis AC, Dickstein DP, Sarfatti SE, Rapoport JL (1997) Implication of right frontostriatal circuitry in response inhibition and attention-deficit/hyperactivity disorder. Journal of the American Academy of Child and Adolescent Psychiatry 36:374-383.

Casey BJ, Cohen JD, O'Craven K, Davidson RJ, Irwin W, Nelson CA, Noll DC, Hu X, Lowe MJ, Rosen BR, Truwitt CL, Turski PA (1998) Reproducibility of fMRI results across four institutions using a spatial working memory task. Neuroimage 8:249-261.

Cummings JL (1995) Anatomic and behavioral aspects of frontal-subcortical circuits. Ann N Y Acad Sci 769:1-13.

Declerck CH, Boone C, Emonds G (2013) When do people cooperate? The neuroeconomics of prosocial decision making. Brain Cogn 81:95-117.

Domes G, Heinrichs M, Glascher J, Buchel C, Braus DF, Herpertz SC (2007) Oxytocin attenuates amygdala responses to emotional faces regardless of valence. Biol Psychiatry 62:1187-1190.

Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis. II: Inflation,

flattening, and a surface-based coordinate system. Neuroimage 9:195-207.

Floresco SB, St Onge JR, Ghods-Sharifi S, Winstanley CA (2008) Cortico-limbic-striatal circuits subserving different forms of cost-benefit decision making. Cogn Affect Behav Neurosci 8:375-389.

Giedd JN (2004) Structural magnetic resonance imaging of the adolescent brain. Ann N Y Acad Sci 1021:77-85.

Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, Paus T, Evans AC, Rapoport JL (1999) Brain development during childhood and adolescence: a longitudinal MRI study. Nature Neuroscience 2:861-863.

Gogtay N, Giedd JN, Lusk L, Hayashi KM, Greenstein D, Vaituzis AC, Nugent TF, 3rd, Herman DH, Clasen LS, Toga AW, Rapoport JL, Thompson PM (2004) Dynamic mapping of human cortical development during childhood through early adulthood. Proc Natl Acad Sci U S A 101:8174-8179.

Goldberg, II, Harel M, Malach R (2006) When the brain loses its self: prefrontal inactivation during sensorimotor processing. Neuron 50:329-339.

Gupta R, Koscik TR, Bechara A, Tranel D (2011) The amygdala and decision-making. Neuropsychologia 49:760-766.

Halperin JM, Schulz KP (2006) Revisiting the role of the prefrontal cortex in the pathophysiology of attention-deficit/hyperactivity disorder. Psychol Bull 132:560-581.

King-Casas B, Chiu PH (2012) Understanding interpersonal function in psychiatric illness through multiplayer economic games. Biol Psychiatry 72:119-125.

Kirsch P, Esslinger C, Chen Q, Mier D, Lis S, Siddhanti S, Gruppe H, Mattay VS, Gallhofer B, Meyer-Lindenberg A (2005) Oxytocin modulates neural circuitry for social cognition and fear in humans. Journal of Neuroscience 25:11489-11493.

Koscik TR, Tranel D (2011) The human amygdala is necessary for developing and expressing normal interpersonal trust. Neuropsychologia 49:602-611.

Lichtman JW, Colman H (2000) Synapse elimination and indelible memory. Neuron 25:269-278.

Luna B, Sweeney JA (2004) The emergence of collaborative brain function: FMRI studies of the development of response inhibition. Ann N Y Acad Sci 1021:296-309.

Moriguchi Y, Ohnishi T, Mori T, Matsuda H, Komaki G (2007) Changes of brain activity in the neural substrates for theory of mind during childhood and adolescence.

Psychiatry Clin Neurosci 61:355-363.

Parent A, Hazrati LN (1995) Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. Brain Res Brain Res Rev 20:91-127.

Rakic S, Zecevic N (2000) Programmed cell death in the developing human telencephalon. EurJNeurosci 12:2721-2734.

Rilling JK, Sanfey AG (2011) The neuroscience of social decision-making. Annu Rev Psychol 62:23-48.

Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD (2004) The neural correlates of theory of mind within interpersonal interactions. Neuroimage 22:1694-1703.

Rubia K, Smith AB, Woolley J, Nosarti C, Heyman I, Taylor E, Brammer M (2006) Progressive increase of frontostriatal brain activation from childhood to adulthood during event-related tasks of cognitive control. Hum Brain Mapp 27:973-993.

Schepis TS, Adinoff B, Rao U (2008) Neurobiological processes in adolescent addictive disorders. Am J Addict 17:6-23.

Shaw P, Kabani NJ, Lerch JP, Eckstrand K, Lenroot R, Gogtay N, Greenstein D, Clasen L, Evans A, Rapoport JL, Giedd JN, Wise SP (2008) Neurodevelopmental trajectories of the human cerebral cortex. Journal of Neuroscience 28:3586-3594.

Sowell ER, Thompson PM, Holmes CJ, Batth R, Jernigan TL, Toga AW (1999) Localizing age-related changes in brain structure between childhood and adolescence using statistical parametric mapping. Neuroimage 9:587-597.

Sowell ER, Peterson BS, Thompson PM, Welcome SE, Henkenius AL, Toga AW (2003) Mapping cortical change across the human life span. Nature Neuroscience 6:309-315.

Steinbeis N, Bernhardt BC, Singer T (2012) Impulse control and underlying functions of the left DLPFC mediate age-related and age-independent individual differences in strategic social behavior. Neuron 73:1040-1051.

Tekin S, Cummings JL (2002) Frontal-subcortical neuronal circuits and clinical neuropsychiatry: an update. J Psychosom Res 53:647-654.

Tye KM, Prakash R, Kim SY, Fenno LE, Grosenick L, Zarabi H, Thompson KR, Gradinaru V, Ramakrishnan C, Deisseroth K (2011) Amygdala circuitry mediating reversible and bidirectional control of anxiety. Nature 471:358-362.

van't Wout M, Sanfey AG (2008) Friend or foe: The effect of implicit trustworthiness

judgments in social decision-making. Cognition 108:796-803.

van Honk J, Eisenegger C, Terburg D, Stein DJ, Morgan B (2013) Generous economic investments after basolateral amygdala damage. Proceedings of the National Academy of Sciences.

Westlye LT, Walhovd KB, Dale AM, Bjornerud A, Due-Tonnessen P, Engvig A, Grydeland H, Tamnes CK, Ostby Y, Fjell AM (2010) Differentiating maturational and aging-related changes of the cerebral cortex by use of thickness and signal intensity. Neuroimage 52:172-185.

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.

Yang Y, Raine A (2009) Prefrontal structural and functional brain imaging findings in antisocial, violent, and psychopathic individuals: a meta-analysis. Psychiatry Res 174:81-88.

6 Valuing Others: Evidence from Economics, Developmental Psychology, and Neurobiology

Book chapter published in *Neuroscience and Social Science: The Missing Link* (pp. 21-45. Springer, Cham.) by:

Billeke, P.[1], Soto-Icaza, P.[2], **Aspé-Sánchez, M.**[1], Villarroel, V.[3], & Rodríguez-Sickert, C.[1]

[1] División de Neurociencias (NeuroCICS), Centro de Investigación en Complejidad Social, Facultad de Gobierno, Universidad del Desarrollo, Santiago, Chile.

[2] Interdisciplinary Center of Neuroscience, Pontificia Universidad Católica de Chile, Santiago, Chile.

[3] Centro de Investigación y Mejoramiento de la Educación (CIME), Facultad de Psicología, Universidad del Desarrollo, Concepción, Chile.

6.1 Abstract

Human social skills are widely studied among very different disciplines. In this chapter, we review, discuss, and relate evidence concerning the process of valuing others' perspectives, preferences, and behaviors from an economic, psychological, and neurobiological viewpoint. This process of valuing others (or other-regarding preferences) can be understood as weighing others' preferences to adapt our own behavior and achieve adequate social interaction. We first review economic research related to decision-making in social contexts, with emphasis on how decision-making has integrated other-regarding preferences into the decision-making algorithm. By means of social and developmental psychology research, we then review how social skills develop from identification to understanding others. Finally, we discuss the neurobiological mechanisms underlying social skills and social decision-making, focusing on those systems that can participate in processes of valuing others preferences. As a conclusion, we highlight five points that we believe an interdisciplinary approach should take into account. We thus intend to generate a starting point for building a more extensive explicatory bridge among the different disciplines that study complex human social behavior.

Keywords: Neuroeconomics, Decision-making, Other-regarding preferences, Mentalization, Theory of mind, Social cognition, Interdisciplinary approach, Game theory.

6.2 Introduction

We are an extremely social species; almost all of our behavior is related to other human beings. Currently, various disciplines deal with the problem of understanding human social behavior. However, few proposals that combine different approaches and findings have been elaborated. In this chapter, we discuss the evidence and research approaches from an array of disciplines related to the idea of how humans consider other preferences and behaviors during this decision-making process. We shall use the term "valuing others" to refer to the processes by which humans weigh the preferences and behaviors of others as to adapt or guide their behavior during social interactions. In the following pages, our endeavor will be to present and discuss the evidence from three research programs, namely, (1) economics research related to decision-making in social contexts, (2) social psychology research related to the development of mentalizing and perspective-taking skills, and (3) neuroscience research related to neuronal mechanisms underlying vicarious human behaviors.

The fundamental aim of this chapter is to show some of the current efforts to build an interdisciplinary understanding of social behavior instead of giving a global integrative approach. In order to build a fully interdisciplinary research programming between social science and neuroscience, the authors have established some basic bridges which are necessary to discuss and begin to build this understanding. Therefore, with the purpose of contributing to this global aim, we have structured this chapter in three sections. In the first one, we discuss how the approach from economics toward the social decision-making process has started to incorporate social preferences and how neuroscience approaches can contribute to improving the predictive ability of the behavioral model. In the second section, we review evidence from developmental psychology related to how human beings begin to understand

and integrate the perspective of others into their own behavior and decisions. Finally, we discuss findings from social neuroscience and neuroeconomics related to the neurobiological mechanisms that underlie social decision-making, in order to suggest possible interdisciplinary approaches, and their possible pitfalls.

6.3 Behavioral Models of Human Conduct and the Black Box

In recent years, the emergence of subfields such as neuroeconomics and social neuroscience has driven the dialogue between behavioral economics and natural science. Especially, behavioral economics has relied on game theory an experimental paradigm for neuroscientists when studying complex social behavior inside the controlled settings of a laboratory. Likewise, to concurrently record or modulate brain activity—by means of techniques such as electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and transcranial magnetic stimulation (TMS) (see below) – could shed a light on the cognitive mechanisms that underlie the behavior of experimental subjects and their reactions against the behavior of their fellow partners.

When there is a confluence of disciplines, the potential gains of combining both perspectives might be hampered by language barriers (e.g., jargon that is discipline specific) and incongruities between the widespread research practices within each discipline (e.g., the importance that is given to generality in contrast with parsimony or to prediction over explanation). In this section, we suggest three perspectives that can lead to a fruitful interdisciplinary interaction from the perspective of economics. We focus on (1) the neurophysiological foundations of behavioral models of social preferences, (2) general guidelines for modeling social behavior and social cognition,

and (3) specific instantiations of neurophysiological variables within those behavioral models.

6.3.1 *Homo behavioralis* and the Influx of Ideas from Psychology and Other Disciplines

When scholars from disciplines such as psychology or anthropology began to question the plausibility of the prevalent model of human agency in economics, the reply came from one of his most renowned representatives. Milton Friedman wrote his famous Essays in Positive Economics (1951), which strongly influenced future generations of economist researchers [1]. There he claimed that "the only relevant test of the validity of a hypothesis is the comparison of its predictions with experience." Furthermore, Friedman argued that even if assumptions appear false or implausible, their empirical weakness should be tolerated if they lead to accurate predictions.

When Friedman adds this second statement, not only can one infer that he was oblivious to the advances of neuroscience but also that the aim of Friedman and his fellow custodians was to keep the black box closed and to keep the *Homo economicus* locked inside [2]. It is not that the members of the congregation for the Doctrine of the Economic Faith denied the existence of other drivers of human behavior beyond self-interest –e.g., altruism. Nor did they believe we are perfect optimizers. Their stance relied on an argument of parsimony: the benefits of generalizing the utility function to account for possible anomalies and produce more accurate predictions would be negligible against the loss of parsimony and tractability of adding new parameters to the utility function. The overwhelming amount evidence from laboratory and field experiments showed that this view on the trade-off between prediction power and parsimony was not accurate. The effort to correct this mistake was as-

sumed by a new breed of "behavioral" economists. Indeed, one can say that there is nothing new in this approach. They are just continuing the enterprise launched by Adam Smith himself, as a moral philosopher, in The Theory of Moral Sentiments (1759) [3].

The first task undertaken by the behavioral squad was to upgrade the utility function so that these "anomalies" could be captured within an augmented utility function. Around psychological constructs, such as loss aversion and reference dependence, Kahneman and Tversky developed prospect theory [4]. While Kahneman, Tversky, and their followers focused on decision under uncertainty, and later on issues such as intertemporal inconsistency [5, 6], a separate group of behavioral economists reacted to the strong evidence against the self-interest hypothesis provided by experimental studies. These studies showed that agents do cooperate in social dilemmas such as trust games [7–11], public good games [12–14], even when cooperating is against their (material) self-interest. And, within bargaining games such as the ultimatum game [15–17], agents are willing to incur in material costs to avoid unfair outcomes and sanction free riders in collective action problems [18].

Taking their insights from social psychology, sociology, and anthropology, a family of models was produced within behavioral economics. These models, referred to as models of social (or other-regarding) preferences, can be either outcome based, e.g., models of inequity aversion [19, 20], or intention based, e.g., models that capture norms of both positive and negative reciprocity [21–23]. Cooperation in trust games was initially understood as the result of positive reciprocity (intention-based social preferences). The trustee is willing to spend resources to reward trust placed in him. On the other hand, rejection on the ultimatum game was initially understood as the result of inequity aversion. However, later studies provided evidence for a more

complex structure of moral response. Trustees in a trust game are also motivated by outcome-based preferences [8], and rejection in the ultimatum game also involves negative reciprocity [24]. Furthermore, current studies show that the research on social preferences can also be extracted by the research produced in other areas of behavioral economics. For instance, time inconsistency can also affect the nature of social preferences [25].

To the extent that neuroeconomic studies have provided neurophysiological mechanisms for experimental anomalies and, thus, biological foundations for social preferences models, neuroeconomists were welcomed as part of the new tribe of behavioral economists but were not so well received by old-school orthodox economists who were still concerned with keeping the black box closed even for the new model of human agency: the homo behavioralis and its representation in an augmented utility function. For instance, it has been argued that neuroscience could not transform economics because what goes on inside the brain is irrelevant to the discipline. As if nothing had changed since Friedman's influential piece, they put forward the idea of a "mindless economics," arguing that what matters are the decisions people make, not the process by which they reach them [25]. We will develop this idea in the opposite direction and claim that the major challenges posited by neuroeconomics precisely relate to our understanding of the neurocognitive processes that underlie social behavior and, furthermore, open the possibility to embed economics in the biological processes taking place in the brain.

6.3.2 Impact of Neurosciences on Modeling Individual and Social Behavior

In the same way that behavioral economics has used insights from psychology to develop more "realistic" models of individual decision-making, in which people often did things that were not in their best interests, the evidence coming from neurobiology presents an additional challenge to the standard economic assumptions. Thus, evidence from neuroeconomics indicates that decision-making is far from being a unitary process (a simple matter of integrated and coherent utility maximization), suggesting instead that it is driven by the interaction of multiple systems or processes [26]. This range from the more basic dual-process approach that has influenced our general comprehension of human cognition and behavior beyond Descartes' error (fast/hot module and the slow/cold, automatic vs. controlled pro- cesses [26–28]) to more complex multiple system approaches toward social behavior and social decision-making [29–31]. Steinbeis et al. [32], for instance, show that behavioral inhibition—modulated by the neuroanatomical development of the cognitive control systems—plays a crucial role in the implementation of fair behavior in bargaining games.

6.3.3 Prediction Accuracy of Behavioral Models: Combining Psychological and Neurobiological Parameters

A specific aspect of the relevance of the neuroeconomic program refers to its capacity to inform behavioral models in such a way that prediction accuracy can be improved. This point is very important, because if we do not build a bridge between neuroscience and algorithmic social decision theory, it will be very difficult for this program to reach the academic community of economists. To discuss the issues that

could emerge from this challenge, we consider a distributional problem in the spirit of Andreoni and Miller [33], in which an agent i decides how to split an amount m between himself and another agent $_{-i}$ for different budget constraints. For every monetary unit agent i sacrifices $(m-x_i)$, his partner will receive $(m-x_i)/p$ monetary units. Thus, p can be interpreted as the price of altruism and agent i's choice can be represented as the consumer's choice problem.

Neoclassical Model (*Homo economicus*, Black Box)

$$max_{x_i} \{U_i(x_i)\} \, s.t. x_i + px_{-i} = m \tag{14}$$

which yields to $x_i(m,p) = x_i^*$ with $x_i^* = m$ and $x_{-i}^* = 0$.

In the case above, the only relevant argument of $U_i(.)$ is his own material self-interest x_i. If, alternatively, we consider that agent i's choice is also affected by the material welfare of his partner $-i$, we could represent his choice problem as follows:

Behavioral Model (Other-Regarding Preferences, Black Box)

$$max_{x_i,x_{-i}} \{U_i(x_i, x_{-i}, \theta_i)\} \, s.t. x_i + px_{-i} = m \tag{15}$$

which yields to $x_i(m,p,\theta_i) = x_i^*$ with $x_i^* = m$ and $x_{-i}^* = 0$, where θ_i is a parameter that represents the intensity of the moral dispositions of the agent that counterbalances his self-interest. Most models assume that θ_i is private. Now consider the possibility that θ_i can be estimated from the neurobiological activation n_i, $\hat{\theta}_i(n_i) = \theta_i + \epsilon_i$. If this is the case, the lower the measurement error, the greater will be the predictive gains of opening the black box. The registered neurobiological

activation n_i could give us information about θ_i through two channels: the individual's idiosyncratic characteristics and the dimensions of the stimuli not captured by the model. For the sake of simplicity, we will assume that n_i is simply a contextual modulator of θ_i. Thus, the structure of choice could be represented as follows.

Neurobiological Model (Other-Regarding Preferences, Neurobiological State)

$$max_{x_i,x_{-i}} \{U_i(x_i, x_{-i}, \theta_i|n_i)s_i\} \; s.t. x_i + px_{-i} = m \tag{16}$$

which yields to $x_i(m, p, \theta_i|n_i) = x_i^{**}; x_i(m, p) = x_i^{**} > 0, x_{-i}^{**} > 0$. The improvement in prediction accuracy of a model that incorporates n_i is an indicator of the incompleteness of the behavioral model. However, it is not only important to come up with a model that accurately predicts behavior in a particular context. Fehr and Camerer [34], among others, argue that prosocial behaviors occur in one-shot anonymous games as the result of a reflexive behavior that is highly adapted for repeated interactions where immediate prosocial behavior earns future benefits. Under this view, prosociality in one-shot games results from bounds on rationality in full response to changes in the economic structure. Alternatively, prosocial behavior could reflect robust social preferences for treating others generously or reciprocally, and those preferences are similar to preferences for other kinds of primary and secondary rewards. Within this scheme, different arrangements of neurobiological activation $n_i^0; n_i^1$ could lead to similar predictions in terms of cooperation that could indicate the motives underlying both cases. Such a case has been shown recently; see below [35]. Furthermore, these neural traits could provide crucial information to distinguish different types of individuals and, consequently, have more information about their behavior in the future or in different social contexts.

A crucial issue in this point is what are precisely these neurobiological traits and states and how these states weigh the parameters of self-interest and other-regarding preferences. Although neuroscientists are far from reaching consensus, there is accumulative evidence that can indicate some general structures of these traits and states. In the following section, we will review some critical evidence from developmental psychology and developmental neuroscience in order to give insight on how these neurobiological states mature and change during the ontogeny. Then, in the final section, we shall analyze how these neurobiological processes can be structured, with special focus on how the system weighs and values the regarding preferences of others.

6.4 Development of Social Preferences

One of the most relevant facts indicates that the neurobiological state has a decisive influence in the decision-making process is the human development. The maturity of different brain systems in different timelines generates several behavioral manifestations that are characteristic to a specific age [36, 37]. This is true not only during childhood and adolescence but also for older adults where pathological neuronal degeneration is expected [38, 39].

Regarding early human development research, one of the most intriguing human social phenomena is the ability to read the minds of others, known as "mentalization" or "theory of mind". This ability has been described as one of the major landmarks in social development, because it enables children to handle more complex social interactions. Indeed, the ability to figure out and finally to attribute and understand the other person's thoughts and feelings has been depicted as a distinctive human trait [40]. However, the mechanism by which this ability emerged has been the sub-

ject of drawn-out controversy [41–43]. The analysis of the development of human social functioning is a useful tool for understanding how social skills are structured. This analysis reveals that social ability development is not a unitary or an "all-or-nothing" type of outcome. Instead, it is an interactive specialization that entails both the association of an ability with a brain system and the specialization of this function in interaction with others [44]. In this context, one of the main drivers for this development is the necessity to anticipate and predict the behavior of others, which is crucial for both primate and human survival [45]. Certainly, the newborn ability to discriminate a relevant biological agent seems to be coordinated to, first, a guarantee that the partner is actually a living being and, second, that this living being is actually human. As human babies are born premature [43], their extreme dependency puts them at higher risk; hence, they must draw the interlocutor's attention directly to them in order to modify the performance of others to get what he or she needs to survive. It seems possible that the later human ability to "read minds" arises from all those previous early stages of social development as a guarantee for survival since it constitutes a specialized expertise of social prediction. This section is organized in three overlapping stages of development, starting with the early capacity to identify biological/social agents and ending with the explicit manifestation of mentalization skills.

6.4.1 Identification of Social Agents in Newborns and Infants

The early stages of social development are the building blocks in which further social skills have grown. Certainly, the only way that a human infant can survive is if there is another being that can provide food, water, etc. Evidence in newborns showed that toddlers as young as only few days of age are able to discriminate different perceptual

signs that indicate the existence of a social agent [46–48]. For example, they can identify points that emulate a coherent biological motion [48], face-like patterns [49], and direct versus averted gaze in faces [50, 51], and they can even imitate basic movements from another human being [52]. Indeed, from 2 months old, infants show a preference for looking at eyes rather than mouths or bodies [53]. This preference also describes a specialization process in 3-month-old toddlers, who prefer eyes only when they are accurately located in the upper part of the face configuration rather than placed in another location of the face [54]. All these findings are showing that there is an ontogenetic orientation toward the social agents, which seems to be in a growing process of behavioral and neural specialization. Indeed, comparative studies between preterm and full-term infants and among subjects of different ages [55–58] emphasize the role of the experience in the cerebral functions refinement [44]. From biological motion detection to imitation and face-like stimuli and direct gaze preference in newborns, human social development seems to be organized to detect, understand, and finally predict and manipulate the social agent [59].

EEG findings in infants and children are in accordance with this developmental perspective. The EEG technique is a noninvasive measurement of the brain activity through scalp electrodes widely used in neuroscience [60]. The evidence has shown that the electrical brain activity phase related to stimulus presentation, called event-related potentials (ERPs) [61], follows a developmental trajectory. An illustrative example is the N170, that is a negative deflection occurring at 170 ms after presentation of a human [60, 62–69], whose likely source is the ventral visual stream near the fusiform face area. In adults, the N170 evidenced a higher amplitude and latency for inverted human faces, while in infants it did not show any modulation by stimuli orientation. In 6-month-old infants, there is an "infant N170" (called P400

component) characterized by higher amplitude in response to faces displaying direct gaze rather than an averted gaze [50], as well as to inverted faces only in the case of their mother [70], evidencing a specific selection process present in early life.

6.4.2 Being Able to Interact with Other Humans

It is important to note that these skills are present in a context of reciprocal interaction [71, 72]. While it is clear that infant behaviors like crying, screaming, gazing, and smiling are aimed to make the social partner answer their requirements, it is also clear that the partner cannot remain indifferent to those calls of attention. What actually happens when infants and their caregivers are coordinated or synchronized? It has been described that in mother-toddler relationships with infants from 3 to 6 months, the engagement periods came in a burst mode, with periods of asynchronous states [72]. Interestingly, these mismatch states are followed by repair sequences of the interactive errors by both the infant and the mother. These repair behaviors can have functionality in the interaction skills development. Indeed, the importance of stages as "reparation" contexts has been widely described in the attachment theory [71]. Precisely, these bonding-recovering stages emphasize the importance of the mutuality of the attachment between the caregiver and the infant which is crucial to underline [71, 73]. The higher social skills like mentalization abilities were the result of all these precursors or early stages of development, which are the building blocks in which further social skills are grown [59].

An important step in the development of the capacity to interact with other human beings is the joint attention (JA) skill. JA has been described as the capacity to share an interest with another person by alternating the gaze in order to coordinate the interest in an object with a social partner [74–80]. A key component of JA is the

division and the alternation of the subject's attention between the object and the partner [77, 81]. Several studies agree that JA emerges around the age of 9 months [74, 76, 77, 82], when children learn to use eye contact to derive information about another person's goal-directed behavior [76]. Importantly, the ability to attend to an object jointly with another person has proved to be crucial for several capacities such as social synchronization, development of language [74, 76, 78, 79, 83, 84], and development of theory of mind [80]. The knowledge of the latter tends to be ambiguous to clarify if JA involves a level of "self-awareness" of the social agent [45]. Does the infant actually "know" the agent's state of mind when is engaged in a JA interaction? There is a line of studies that defines JA as the situation in which two subjects are looking at the same object but without the awareness that the focus of attention is a common interest. The real capacity to realize that the focus of attention is a common element between the infant and the agent is what is called "shared attention" [45]. Accordingly, what is clearly a higher development of social knowledge is the mentalization ability, which is the capacity to understand and predict the behavior of other people and their knowledge, intentions, emotions, and beliefs [85, 86]. Furthermore, JA and shared attention would be intermediate stages toward mentalization inasmuch as the theory of mind ability solely enables to notice and take into account the agent's mental state. Interestingly, the neuroimaging evidence revealed that JA and mentalization might be related. Specifically, fMRI is a method that measures changes in the hemodynamic brain response associated with neural activity—specifically, the blood-oxygen-level-dependent (BOLD) signal [87]. There is broad consensus about the brain network that is recruited when adult subjects participate in mentalizing tasks (see next section below). Interestingly, the same network is involved when participants show JA behavior in adulthood and later

childhood. During early childhood, the EEG evidence shows that responses to JA are associated with the Nc component. This ERP refers to a negative deflection that occurs around 300–850 ms after stimulus onset [56, 66, 77, 82], and it is associated with attentional reorientation. In children during the age when they can achieve the false-belief mentalization, this component did not seem to present any differences. However, two neuronal measures seem to mark the mentalization achievement. One of these is the presence of a specific oscillatory activity in the temporoparietal areas of the mentalization network (see next section) and the maturation on neural fiber that connects the frontal and temporoparietal regions [88]. Thus, specific neuronal development seems to be a marker for more complex social skills achievement.

6.4.3 Knowing the Others' Mental States

What do infants know about the mental states of others? Do they actually try to modify the actions of others because they can infer what is in their minds? Premack and Woodruff [89] stated that the mentalization ability is a system of inference that enables us to attribute mental states both to oneself and to another—for instance, purposes, intentions, knowledge, belief, and thinking. Certainly, this system of inference is needed because such "mental states" are not directly observable, making it a "theory" of what are the others' mental states (i.e., theory of mind). The explicit skill to identify other people's false beliefs becomes evident not before 4 years of age [85, 90]. However, there is a line of research that describes how infants are able to do some kind of inferences about others' feelings and thoughts [91–94]. That line of studies appeared as alternative experimental paradigms to overcome the language-dependent bias which standard/classic false-belief tasks [86] have. Hence, the infants' difficulty both to inhibit their own knowledge about something that an-

other person does not know and to think over different representations makes this task impossible to solve for children under 4 years old [95, 96]. Therefore, researchers use infants' longer looking time as a measure of children's anticipatory belief [94] or surprise as measure of a violation of the expectation paradigm [92, 93, 97] in non-verbal false-belief tasks. Thus, this line of research has shown that there is evidence of an "implicit" theory of mind [91]. However, there is another line of research that has been skeptical about this interpretation [41–43, 98, 99]. This evidence can be interpreted just as perceptual processes and competences rather than high-level cognitive processes. Furthermore, high-level constructs that come from this experimental paradigm might be revealing the researcher's over-interpretation instead of the ability for which it was created [43]. Indeed, the increase in looking times that these studies have shown might be revealing a visual perception process related with a new arrangement of the stimuli rather than an interpretation of the agent's belief [99].

At this point of the controversy, it is important to consider that the implicit mentalization ability, the JA ability, the different levels of visual perspective taking (mentioned below), and the explicit theory of mind itself could be understood as stages of complexity inside the development process of the same capacity. The visual perspective taking (VPT) is the capacity to know that an object can be seen from a certain point of view and that someone else could not see it because there is a physical barrier [100]. Research of VPT should also be considered to understand the mentalization development as a dynamic building block process. These studies provide interesting evidence to consider the existence of an intermediate level of mentalization [59]. The first level of VPT [101, 102] can be understood as a previous step toward a well-consolidated theory of mind, because, around the age of 2, the

child is only able to identify whether another person can see an object or not, but it says nothing about a genuine capacity to attribute the mental state of the agent. Nevertheless, this VPT level becomes more complex a couple of years after when it allows the child to identify the others' references and perspectives [90, 98, 101, 102]. This higher VPT level, known as Level 2 VPT, allows the child to understand that objects can be seen in different ways, depending on the form of presentation and point of view [98, 101, 102]. There is evidence that correlates Level 2 VPT with the development of mentalization ability [101]. Although the first theories point out that the visual perspective taking is the basic process from which more complex (social) perspectives arise, recent evidence indicates an opposite ontological development [102]. Early infants can track others' experiential backgrounds. In fact, several studies have found that infants take what others have witnessed into account when acting and responding toward them. In other words, the infants revert to the background constituted by past experiences and use it to understand an agent's desires, goals, and intentions. This ability becomes evident before infants can solve complex visual perspective-taking tasks (Level 2) and even before they can solve explicit mentalizing problems, like the false-belief task [91]. This evidence indicates that the developmental processes that lead to the explicit mentalizing ability are related to the integration of others' preferences into our behaviors. This skill, as an integrative process, becomes more complex through aging, incorporating more sources of information, such as memories, social knowledge, and visual skills, among others. Thus, the development of this skill serves as the basis for more complex explicit mentalizing or the theory of mind skill.

Following the deconstruction of the mentalizing concept proposed elsewhere [103], the skill of valuing others can help us gather not well matching evidence, which has

come from cognitive neuroscience and neuroeconomics. In the next section, we will review neuroscience evidence related to brain components of the system of other-regarding preferences.32

6.5 Neurobiological System Related to Other-Regarding Preferences

Our brain has evolved to solve complex cognitive demands required for living in social groups of increasing size [104]. Experimental evidence has established that, unlike other social species, humans display a large amount of cooperative behaviors, including altruism, trust, and reciprocity [105, 106]. These behaviors are observed even when individuals interact with strangers and with individuals they will never meet again [107]. Trust, altruism, and reciprocity are crucial to establish and maintain cooperative links between different individuals. Recent work using neuroscience techniques has begun to reveal the brain states related to these prosocial dispositions [108]. In the following subsection, we will review evidence from neuroeconomic studies using two game theory experimental paradigms, namely, trust and dictator games. Then, we shall discuss evidence from the two putative systems related to other-regarding preferences or "valuing others" processes that can underlie human prosocial behaviors.

6.5.1 Trust and Reciprocity

The most widely used experimental setting to study trust and reciprocity is the trust game (TG) or invested game. In this game, two players, who do not know each other, engage in an anonymous interaction. The experimenter gives the "investor" (or trustor) some amount T of money. The trustor then decides how much of T

send (or "invest") in the other player, referred to as the trustee. The amount A_1 sent by the trustor is multiplied by an exchange factor r (typically 3). Thus, the trustee receives an amount of money three times the amount sent by the trustor (rA_1). Finally, the trustee decides how much of the money received (rA_1) is sent back to the trustor (A_2) [7]. The prediction from the self-interest hypothesis for TG is that the trustees will keep all the money. Assuming that the trustors have mentalizing capabilities (see above), they should anticipate this betrayal and send nothing. In the very first test of this game, 0.6% of the trustors sent nothing to the trustee, 66% sent half or more of their endowment, and about 50% ended the game with more money than their initial endowment (which implies, of course, that $A_2 > A_1$; in other words, trustees were trustworthy [7].

These behaviors have been replicated in several studies. In a recent meta-analysis, Johnson and Mislim [109] collected the data from the 162 replications of the TG available at the time and found that, on average, trustors send 0.5 of his/her endowment to the trustee (n = 23,900; std = 0.12; min = 0.22; max = 0.89), while the trustee returns 0.37 of their total endowment (n = 21,529; std = 0.11; min = 0.11; max = 0.81 [109]. Repeated interactions of the TG show a similar pattern, indicating a high tendency toward trust and reciprocity by both players [107].

Trustee behavior is interesting. While, for trustors, there is an expected gain, this is not so clear for trustees. The trustee has the opportunity to break the trust, which is, as stated above, the classical self-interest prediction. This is particularly true for one-shot, anonymous interactions since there are no incentives to build reputation and create a greater amount of trust for future interactions. Classically, trustee's behavior has been considered just reciprocity, but this is only true if allocations made by trustees are different from allocations made by a subject in a context where

his/her behavior is unrelated with the perceived intentions of cooperation from the other player [110]. There is a difference between intention-based behaviors, such as the behavior in the TG, where trustee's behavior depends on ascribing cooperative intentions to the trustor, and outcome-based behaviors, such as the behavior in the Dictator game (DG, described below), where subject behavior depends only on the final share of the game and not on the others' intentions.

6.5.2 The Neural Dynamics of the TG

In a TG, the very first decision by the trustor involves deciding whether to trust the other player or not. From the trustors' perspective, this involves (1) knowing whether they are playing with another human or a non-intentional entity (generally a computer which makes random allocations) and (2) then deciding to send or not to send some amount of money to the trustee. Several reports have shown increased activity in the medial prefrontal cortex (mPFC, a structure involved in metallization processes; [111]) when trustors decide to trust another human partner [112–115]. In addition, during the first stage, the trustor has not received any feedback on the trustworthiness of his/her partner; therefore, the reinforcement learning system must be engaged to adjust trustor behavior based on feedback reward. Delgado et al. [114] read the descriptions of the life events of different trustees to trustors, indicating praiseworthy, neutral, or suspicious moral characters for each of them. Not surprisingly, rates of cooperation were higher when playing with the praiseworthy partner. Interestingly, trustors showed different activation in the ventral striatum (VS) for positive and negative feedback but only when they were playing with the neutral trustee. The VS has been involved in processing feedback and prediction error [114, 116], suggesting that, in the neutral condition, trustors activate the reinforcement

system to learn about the trustworthiness of their partners, while praiseworthy and suspicious moral characters bias the behavior of trustors [114].

Interestingly, the neuropeptide oxytocin (OXT) has been associated with trust behaviors in humans [117, 118]. Kosfeld et al. [119] used a TG experiment to show that intranasal infusions of OXT increase trust in humans (but not in other nonsocial interactions), do not increase risk-taking behavior, and did not change trustees' behavior. Although the mechanism of action of OXT is not clear, evidence suggests that OXT decreases stress responses and anxiety in social interactions, likely modulating the amygdala and anterior cingulate cortex (ACC) activity [117, 120]. Considering now the situation of the trustees, reports show that the mentalization system becomes active when they receive an allocation from trustors. Van den Bos et al. [121] has shown that the mPFC increases its activation when trustees defect. On the other hand, when trustees reciprocate a high-risk allocation (i.e., the trustor could lose a large amount of money if the trustee chose to defect), there is greater activation of the temporoparietal junction, which is also a part of the mentalization system [122–124]. Moreover, trustees' reciprocity in low-risk allocations correlated with the activity in the anterior insula cortex (AIC), a structure involved in emotional and salience processing [113, 125]. Furthermore, trustees reciprocating low benefit allocations (i.e., when the monetary incentives to reciprocate are low) were associated with an increased activity in the ACC and the dorsolateral prefrontal cortex (dlPFC), which are structures involved in cognitive control and the inhibition of selfish impulses [126–129].

Another interesting finding is the effect of individual traits in reciprocal interaction [121, 130]. For example, people with more traits characterized by positive emotionality trust more in others, while people with less tendency to psychopathic traits show

more reciprocate behaviors [130]. Other study shows that when a prosocial subject reciprocated, they showed an increased activation in VS, while defection increased the activity in ACC, AIC, and right TPJ. In contrast, pro-self individuals showed the opposite pattern, showing increasing ACC, AIC, and right TPJ activity after they reciprocated. This shows that these structures were more active when participants chose their less frequent behavior, considering their personal trait or past history [121].

Trustees' reciprocal behavior is also influenced by expectations [131]. Chang et al. [131] asked trustees about their second-order beliefs (i.e., how much money they think the trustor expects) and compared these second-order beliefs with the amount that trustees actually send. With this information, they could categorize the allocations made by trustees as "minimizing guilt" (when the amount sent was close to the trustees' second-order beliefs) or "maximizing outcome" (when trustees sent an amount significantly smaller than what they expected based on their second-order beliefs). When trustees minimized guilt, they exhibited higher activation in dlPFC, AIC, and dorsal ACC, which are structures reported to be activated by negative affective states [132–134]. On the other hand, when trustees maximized outcome, higher activation occurred in ventral mPFC, VS, and dorsal mPFC. The authors proposed a model where minimizing guilt increased AIC activation, which increased activation in dorsal mPFC, while maximizing outcome decreased AIC activation, which increased activation in the VS [131].

6.5.3 Altruism

Historically, altruism has been studied by means of the dictator game (DG). In this game, there are also two players involved in an anonymous one-shot interaction. The

first player, called "dictator," receives an amount T of money and donates some a part of it ($A_1 \in [0, T]$) to the second player, called the "recipient." This decision ends the game and the recipient has no participation in deciding about this distribution. Crucially, the recipient has no chances of punishing the dictator if the amount is not acceptable to him. Thus, there are not direct incentives for a strictly self-interested dictator to share any portion of the received money, and any donation is defined as an altruistic act [108, 135].

Behavioral evidence shows that even when participants play this game with unknown others, dictators tend to donate around 25% of their money to the recipient [136]. Interesting variants have been introduced to the game. Cherry et al. [137], for instance, made the dictators earn their own money, thereby giving subjects a sense of ownership. In this case, about 91% of the dictators don't send anything to the recipient. In addition, there have been recent efforts to include social knowledge about the recipient in the DG [138, 139]. Such experimental settings have shown that there are important variables which explain allocations, such as the knowledge about who the recipient is and how the game is explicitly described to the players [136]. Likewise, social distance is an important modulator of behavior in the DG. Hoffman et al. [140] showed that 64% of dictators kept all the money when social distance was maximized. In addition, some authors have shown that donations tend to be higher when people are informed that the recipient is a real charitable organization [138, 139].

6.5.4 Neuronal Dynamic of the DG

Despite its simplicity, and the fact that it has been used widely in behavioral economics, few neuroeconomics experiments have used the DG to assess the neural basis

of altruism. In a recent article, Hutcherson et al. [141] made subjects participate in a DG where subjects had to choose between two options of allocation. By using this protocol, they induced choices between the default 50–50% split, generous (benefiting the other at a cost to oneself) or selfish behavior (benefiting oneself at a cost to another). The authors fitted a drift-diffusion model which assumes that choices are the output of a noisy process that weighs the linear sum of monetary outcomes for self and others. In this model, the choice is made when sufficient neural evidence has accumulated in favor of one of the options, and it assumes that the valuing of self and other outcomes is computed independently and then integrated in an overall value signal. At the neural level, the authors found that ventromedial prefrontal cortex (vmPFC) activity correlated positively with the value that subjects assigned to proposals, as measured by the Likert response scale. vmPFC has been reported to encode stimulus values at the time of decision in a wide range of tasks [142, 143]. Moreover, fitting general linear models (GLM), they found that valuations toward self- outcomes correlate with the activity in both vmPFC and VS, while valuations toward other outcomes correlate with the activity in the right TPJ, precuneus, and vmPFC. These results, further discussed below, show that the right TPJ is an area that becomes activated specifically when focusing on others, while vmPFC combines information about self and others.

In another experiment, Hein et al. [35] studied the role of empathy and reciprocity motives in human altruism. Using a DG, they investigated differences in altruistic behavior from experimental subjects when they observed recipients (1) receiving painful shocks (empathy partner) or (2) giving an amount of money to save some of those empathy partners from painful shocks (reciprocity partner), an action perceived as kind and, thus, one that should elicit reciprocity motives. A baseline

partner neither received painful shock nor was instructed to give money for saving subjects from shock. Authors observed that subjects behave more altruistically toward the empathy and the reciprocity partners, noteworthy, without significant differences in allocations between the two motive inductions. At the neural level, a network consisting of AIC, VS, and ACC was activated in both motive-induction conditions. Moreover, individual pattern of brain connectivity in this network predicts subjects' altruist behavior. Interestingly, this prediction was particular for each treatment. Thus, a positive connectivity between ACC and AIC and a slightly negative connectivity between AIC and VS predict empathy-driven altruism, while a strong bidirectional projection between AIC and ACC and a positive connectivity between AIC and VS predict reciprocity-driven altruism. Additionally, the ACC connectivity to AIC correlates positively with baseline levels of altruism. Notice that, at the behavioral level, both motives were indistinguishable, because motives are a mental construct hidden to revealed preferences. A neuroeconomic approach is able to unravel both motives and their influence on altruistic behavior.

6.5.5 Two Putative Systems for Valuing Others' Outcomes

Anterior Cingulate Cortex and Vicarious Performance Monitoring As seen above, a set of cognitive and affective functions determining the need for adaptive control prove central to economic decision-making [144]. A key neural structure that participates in these functions is the ACC, which is involved in interactions such as reciprocity, choosing the less common behavior [128, 145, 146], empathy and reciprocity-driven motives in human altruism [35], violations of social norms [147, 148], and mediating the effects of OXT in trust behavior [117].

The ACC is the frontal part of the cingulate cortex. Anatomically, the ACC has

classically been subdivided in a rostral (rACC) and a dorsal part (dACC) [149]. The inputs to dACC include the amygdala, AI, orbitofrontal cortex, vmPFC and midbrain, and prominent ventral tegmental area. Its outputs target the lateral PFC, the motor cortex, striatum, subthalamic nucleus, and locus coeruleus [150]. The activity of the dACC has been correlated with almost the whole set of known psychological variables. Broadly speaking, dACC has been considered a key hub in a network of brain regions implicated in domain-general executive functions in humans [127], being important for cognitive control (i.e., our ability to flexibly adjust behavior according to internally maintained goals and away from behaviors that are more automatic but distracted from those goals [149]. Consequentially, there exists some agreement relating the involvement of the dACC in motivation and reward-based decision-making [127, 151].

However, there is no clear consensus on the function of dACC. Currently, two main proposals interpret its functioning: the expected value of control (EVC) theory and the foraging value theory (FVT). EVC [150] proposes that dACC plays a central role in decisions about the allocation of cognitive control based on a cost (for instance, the effort needed) and benefit (for instance, improved performance) analysis that identifies the highest EVC. The FVT theory, on the other hand, argues that difficulty or control allocation is insufficient to account for all dACC activity [152]. Instead, the dACC plays a key role in behavioral flexibility. Its activity reflects the history, weighted by time of occurrence, of previously chosen rewards, computing the value of persisting in the current environment versus the value of switching away from it [153]. Following the evidence review above and other experiments using economic social exchanges [115, 147, 154], some researchers argue that particular areas of ACC track, specifically, behavioral motivation and prediction errors not of self but

specifically of others [149]. In this line, studies suggest that the gyral region of the ACC (ACCg) computes "other-oriented" information (i.e., information about other agents that might be animals or people, rather than ourselves). Apps et al. [155], for instance, examined the brain activity of human subjects when they received cues about the level of an economic reward and the cost incurred for receiving this reward, under conditions in which the costs and rewards pertained to the same experimental subjects or to a third person. In this experiment, ACCg activity correlated with the net value of rewards to be received by the third person when the third person incurred the cost of the effort. By contrast, the ACC sulcus signaled the effort level regardless of whether the effort was exerted by the subject or by a third person [149]. Authors found, "with a striking consistency," that the ACCg responds exclusively to other-oriented information.

Temporoparietal Regions and Valuing Others' Processes

As reviewed above, mentalization is our ability to represent and attribute others' mental or internal states, such as ideas, beliefs, desires, emotions, and motivations [31, 156] Similarly, perspective taking (PT) is the ability to comprehend that the same event or object can be seen or constructed in multiple ways, depending on each subject's point of view. Both processes enable humans to weight others' behaviors and preferences into the subjective valuations that underlie decision-making, a process that can be called "valuing others" [38, 39, 124]. At the neurobiological level, meta-analysis studies have shown that this area becomes active in all the tasks involving PT or mentalization [157]. Furthermore, some scholars have proposed that TPJ is a key neural structure underlying the distinction between self and others' perspectives [156, 158–160]. The involvement of the TPJ in general mentalizing functions can be linked to its anatomical characteristics. TPJ is constituted by the

posterior part of the temporal lobe, the inferior part of parietal lobe, and the lateral part of occipital lobe [161]. This area is a heteromodal association cortex integrating multiple sources of sensory (and non-sensory) information. In addition, this region is located at a maximum synaptic/geodesic distance from sensory and motor areas. This seems to be useful for generating integrative computations addressing inner (abstract) and social processes [162]. There is plenty of evidence highlighting a consistent role of the TPJ in other preferences and how much these preferences affect personal decisions. TPJ is engaged, for instance, when subjects must anticipate others' decisions and behaviors [38, 39, 123, 141, 160, 163], when trustees reciprocate a high-risk allocation when pro-self individuals reciprocate [121], or when dictators evaluate the outcomes of others [141]. All these findings point to the existence of neuronal processes that compute others' preferences and behaviors, where TPJ is a key structure underlying the mechanism that allows us to integrate the others' preferences during a social interaction.

6.6 Conclusions

Currently there is a broad interest to combine evidence from different fields to better understand our complex social behavior. Our review suggests that, while the integration between social and natural sciences is still elusive, the evidence warrants five conclusions that may guide interdisciplinary discussion among behavioral economics, developmental psychology, and neuroscience. In particular, we believe that it is necessary to take care of the following observations:

1. The process of social decision-making can be understood as an algorithmic process that necessarily needs to be in contrast with real decision-making data.

2. In this algorithmic process, humans take into account multiple motivators (parameters), where self-interest (wellbeing/survival) and other-regarding preferences (valuing others' processing) are the most relevant.

3. The ways by which these motivators are finally integrated strongly depend on the neurobiological organization of multiple (not unitary) systems.

4. The neurobiological system (understood as neurophysiological states and traits) implicates both a general and a variable organization.

5. The variations of these neurobiological systems (not only one black box) depend at least on ontogenic (developmental) states, contextual constraints, and individual predispositions.

The social skills analyzed here are only an example of the areas where multiple disciplines have focused their efforts. Currently, it is extremely necessary to work on establishing common concepts in order to gather disperse perspectives. Through this chapter, we intend to generate a conceptual bridge among the knowledge input from psychology, neuroscience, and economics. This is certainly not a global theoretical framework but rather a starting point for building common conceptual framings in order to increase an interdisciplinary dialogue. In this way, we expect to be able to address difficult and unanswered questions about our amazing and, at the same time, conflictive social behavior.

Acknowledgments

This work was supported by Comisión Nacional de Investigación Científica y Tecnológica CONICYT (Grant FONDECYT 11405268 to CR-S, Grant FONDECYT Inicio 11140535 to PB and Grant PCHA/DoctoradoNacional/2014-21140043 to PS-

I).

6.7 References

1. Friedman M. Essays in positive economics. Chicago: University of Chicago Press; 1953.

2. Stigler G, Becker G. De Gustibus Non Est Disputandum. Am Econ Rev. 1977;67(2):76–90.

3. Ashraf N, Camerer CF, Loewenstein G. Adam Smith, behavioral economist. J Econ Perspect. 2005;19(3):131–45.

4. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica JSTOR. 1979;47(2):263–91.

5. O'Donoghue T, Rabin M. Doing it now or later. Am Econ Rev. 1999;89(1):103–24.

6. Laibson D. golden eggs and hyperbolic discounting. Q J Econ. 1997;112(2):443–78.

7. Berg J, Dickhaut J, McCabe K. Trust, reciprocity, and social history. Games Econ Behav. 1995;10(1):122–42.

8. Cox JC. How to identify trust and reciprocity. Games Econ Behav. 2004;46(2):260–81.

9. Falk A, Kosfeld M. The hidden costs of control. Am Econ Rev. 2006;96(5):1611–30.

10. Sheremeta RM, Zhang J. Three-player trust game with insider communication. Econ Inq. 2014;52(2):576–91.

11. Heyes A, List JA. Supply and demand for discrimination: strategic revelation of own characteristics in a trust game. Am Econ Rev. 2016;106(5):319–23.

12. Fehr E, Gächter S. Cooperation and punishment in public goods experiments. Am Econ Rev. 2000;90(4):980–94.

13. Reuben E, Riedl A. Enforcement of contribution norms in public good games with heterogeneous populations. Games Econ Behav. 2013;77(1):122–37.

14. Oprea R, Charness G, Friedman D. Continuous time and communication in a public-goods experiment. J Econ Behav Organ. 2014;108:212–23.

15. Brañas-Garza P, Espín AM, Exadaktylos F, Herrmann B. Fair and unfair punishers coexist in the ultimatum game. Sci Rep. 2015;4(1):6025.

16. Nowak MA. Fairness versus reason in the ultimatum game. Science. 2000;289(5485):1773–5.

17. Güth W, Kocher MG. More than thirty years of ultimatum bargaining experiments: motives, variations, and a survey of the recent literature. J Econ Behav Organ. 2014;108:396–409.

18. Fehr E, Gächter S. Altruistic punishment in humans. Nature. 2002;415(6868):137–40.

19. Bolton GE, Ockenfels A. ERC: a theory of equity, reciprocity, and competition. Am Econ. Rev. 2000;90(1):166–93.

20. Fehr E, Schmidt K. A Theory of fairness, competition and cooperation. Q J Econ. 1999;114(August):817–68.

21. Rabin M. Incorporating fairness into game theory and economics. Am Econ Rev. 1993;83:1281–302.

22. Falk A, Fischbacher U. A theory of reciprocity. Games Econ Behav. 2006;54(2):293–315.

23. Dufwenberg M, Kirchsteiger G. A theory of sequential reciprocity. Games Econ Behav. 2004;47(2):268–98.

24. Falk A, Fehr E, Fischbacher U. On the nature of fair behavior. Econ Inq. 2003;41(1):20–6.

25. Andreoni J, Barton B, Bernheim BD, Aydin D, Naecker J. When fair isn't fair: sophisticated time inconsistency in social preferences. Work Pap. 2016;1996:58.

26. Loewenstein G, Rick S, Cohen JD. Neuroeconomics. Annu Rev Psychol. 2008;59:647–72.

27. Damasio A. Feelings of emotion and the self. Ann N Y Acad Sci. 2003 Oct;1001:253–61.

28. Kahneman D. Thinking, fast and slow. New York: Macmillan and Company; 2011.40

29. King-Casas B, Chiu PH. Understanding interpersonal function in psychiatric illness through multiplayer economic games. Biol Psychiatry. 2012;72(2):119–25.

30. Declerck CH, Boone C, Emonds G. When do people cooperate? The neuroeconomics of prosocial decision making. Brain Cogn. 2013;81(1):95–117.

31. Billeke P, Aboitiz F. Social cognition in schizophrenia: from social stimuli processing to social engagement. Front Psychiatry. 2013;4(February):1–12.

32. Steinbeis N, Bernhardt BC, Singer T. Impulse control and underlying functions of the left DLPFC mediate age-related and age-independent individual differences in strategic social behavior. Neuron. 2012;73(5):1040–51.

33. Andreoni J, Miller J. Giving according to GARP: an experimental test of the consistency of preferences for altruism. Econometrica. 2002;70(2):737–53.

34. Fehr E, Camerer CF. Social neuroeconomics: the neural circuitry of social preferences. Trends Cogn Sci. 2007;11:419–27.

35. Hein G, Morishima Y, Leiberg S, Sul S, Fehr E. The brains functional network architecture reveals human motives. Science. 2016;351(6277):1074–8.

36. McAuliffe K, Blake PR, Steinbeis N, Warneken F. The developmental foundations of human fairness. Nat Hum Behav. 2017;1(2):42.

37. Dalgleish T, Walsh ND, Mobbs D, Schweizer S, van Harmelen A-L, Dunn B, et al. Social pain and social gain in the adolescent brain: a common neural circuitry underlying both positive and negative social evaluation. Sci Rep. 2017;7(February 2016):42010.

38. Ibáñez A, Billeke P, de la Fuente L, Salamone P, García AM, Melloni M. Reply: Towards a neurocomputational account of social dysfunction in neurodegenerative disease. Brain. 2017;140(3):e15.

39. Melloni M, Billeke P, Baez S, Hesse E, de la Fuente L, Forno G, et al. Your perspective and my benefit: multiple lesion models of self-other integration strategies during social bargaining. Brain. 2016;139(11):3022–40.

40. Baars B, Gage N. Social cognition: perceiving the mental states of others. In: Cognition, brain and consciousness: introduction to cognitive neuroscience. 2nd ed. San Diego, CA: Elsevier; 2010.

41. Penn DC, Povinelli DJ. On the lack of evidence that non-human animals possess anything remotely resembling a "theory of mind". Philos Trans R Soc Lond Ser B Biol Sci. 2007;362(January):731–44.

42. Povinelli DJ, Vonk J. Chimpanzee minds: suspiciously human? Trends Cogn Sci. 2003;7(4):157–60.

43. Aboitiz FA. Brain for speech. A view from evolutionary neuroanatomy. London: Palgrave Macmillan; 2017.

44. Johnson MH. Interactive specialization: a domain-general framework for human functional brain development? Dev Cogn Neurosci. 2011;1(1):7–21.

45. Emery NJ. The eyes have it: the neuroethology, function and evolution of social gaze. Neurosci Biobehav Rev. 2000;24(6):581–604.

46. Bertenthal BI, Proffitt DR, Cutting JE. Infant sensitivity to figural coherence in

biomechanical motions. J Exp Child Psychol. 1984;37(2):213–30.

47. Pavlova M, Sokolov A. Orientation specificity in biological motion perception. Percept Psychophys. 2000;62(5):889–99.

48. Simion F, Regolin L, Bulf H. A predisposition for biological motion in the newborn baby. Proc Natl Acad Sci U S A. 2008;105(2):809–13.

49. Macchi Cassia V, Simion F, Umiltaa C. Face preference at birth: the role of an orienting mechanism. Dev Sci. 2001;4(1):101–8.

50. Farroni T, Csibra G, Simion F, Johnson MH. Eye contact detection in humans from birth. Proc Natl Acad Sci. 2002;99(14):9602–5.

51. Farroni T, Mansfield EM, Lai C, Johnson MH. Infants perceiving and acting on the eyes: tests of an evolutionary hypothesis. J Exp Child Psychol. 2003;85(3):199–212.

52. Meltzoff AN, Moore MK. Imitation of facial and manual gestures by human neonates. Published by: American Association for the Advancement of Science Stable. URL: http://www.jstor.org/stable/1744187. 1977;198(4312):75–8.

53. Jones W, Klin A. Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. Nature. 2013;504(7480):427–31.

54. Turati C, Valenza E, Leo I, Simion F. Three-month-olds' visual preference for faces and its underlying visual processing mechanisms. J Exp Child Psychol. 2005;90(3):255–73.

55. Macchi Cassia V, Bulf H, Quadrelli E, Proietti V. Age-related face processing bias in infancy: evidence of perceptual narrowing for adult faces. Dev Psychobiol. 2014;56(2):238–48.

56. Luyster RJ, Powell C, Tager-Flusberg H, C a N. Neural measures of social attention across the first years of life: characterizing typical development and markers of autism risk. Dev Cogn Neurosci. 2014;8:131–43.

57. De Haan M, Johnson MH, Halit H. Development of face-sensitive event-related potentials during infancy. In: De Haan M, editor. Infant EEG and event-related potentials. 1st ed. New York: Psychology Press; 2007.

58. Pena M, Arias D, Dehaene-Lambertz G. Gaze following is accelerated in healthy preterm infants. Psychol Sci. 2014;25(10):1884–92.

59. Soto-Icaza P, Aboitiz F, Billeke P. Development of social skills in children: neural and behavioral evidence for the elaboration of cognitive models. Front Neurosci. 2015;9(September):1–16.

60. Haan M De. Introduction to infant EEG and event-related potentials. In: Haan M, editor. Infant EEG and event-related potentials. New York, USA: Psychology Press Ltd New York; 2002. p. 39–76.

61. Luck SJ. Ten simple rules for designing and interpreting ERP experiments University of Iowa. In: Handy TC, editor. Event related potentials: a methods handbook. Cambridge, MA: MIT Press; 2004.

62. Csibra G, Kushnerenko E, Grossmann T. Electrophysiological methods in studying infant cognitive development. In: Nelson CA, Luciana M, editors. Handbook of developmental cognitive neuroscience. Cambridge, MA: MIT Press; 2008. p. 1–50.

63. Hileman CM, Henderson H, Mundy P, Newell L, Jaime M. Developmental and individual differences on the P1 and N170 ERP components in children with and without autism. Dev Neuropsychol. 2013;36(2):214–36.

64. Itier RJ. N170 or N1? Spatiotemporal differences between object and face processing using ERPs. Cereb Cortex. 2004;14(2):132–42.

65. Courchesne E, Ganz L, Norcia a M. Event-related brain potentials to human faces in infants. Child Dev. 1981;52(3):804–11.

66. Dawson G, Webb SJ, McPartland J. Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. Dev Neuropsychol. 2005;27(3):403–24.

67. de Haan M, CA N. Brain activity differentiates face and object processing in 6-month-old infants. Dev Psychol. 1999;35(4):1113–21. 68. Elsabbagh M, Volein A, Csibra G, Holmboe K, Garwood H, Tucker L, et al. Neural correlates of eye gaze processing in the infant broader autism phenotype. Biol Psychiatry. 2009;65(1):31–8.

69. Johnson MH, Griffin R, Csibra G, Halit H, Farroni T, de Haan M, et al. The emergence of the social brain network: evidence from typical and atypical development. Dev Psychopathol. 2005;17(3):599–619.

70. Balas BJ, Nelson CA, Westerlund A, Vogel-Farley V, Riggins T, Kuefner D. Personal familiarity influences the processing of upright and inverted faces in infants. Front Hum Neurosci. 2010;4(February):1.

71. Bretherton I. The origins of attachment theory: John Bowlby and Mary Ainsworth. Dev Psychol. 1992;28(5):759–75.

72. Tronick EZ, Cohn JF. Infant-mother face-to-face interaction: age and gender differences in coordination and the occurrence of miscoordination. Child Dev. 1989;60(1):85.

73. Harlow HF, Zimmermann RR. Affectional response in the infant monkey: orphaned baby monkeys develop a strong and persistent attachment to inanimate surrogate mothers. Science. 1959;130(3373):421–32.

74. Mundy P, Card J, Fox N. EEG correlates of the development of infant joint attention skills. Dev Psychobiol. 2000;36:325–38.

75. Charman T. Why is joint attention a pivotal skill in autism? Philos Trans R Soc Lond Ser B Biol Sci. 2003;358(January):315–24.

76. Morgan B, Maybery M, Durkin K. Weak central coherence, poor joint attention, and low verbal ability: independent deficits in early autism. Dev Psychol. 2003;39(4):646–56.

77. Striano T, Reid VM, Hoehl S. Neural mechanisms of joint attention in infancy. Eur J Neurosci. 2006;23(10):2819–23.

78. Lachat F, Hugueville L, Lemaréchal J-D, Conty L, George N. Oscillatory brain correlates of live joint attention: a dual-EEG study. Front Hum Neurosci. 2012;6(June):156.

79. Hopkins WD, Taglialatela JP. Initiation of joint attention is associated with morphometric variation in the anterior cingulate cortex of chimpanzees (Pan troglodytes). Am J Primatol. 2013;75(5):441–9.

80. Charman T, Baron-Cohen S, Swettenham J, Baird G, Cox A, Drew A. Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. Cogn Dev. 2000;15(4):481–98.

81. Bakeman R, Adamson LB. Coordinating attention to people and objects in mother-infant and peer-infant interaction. Child Dev. 1984;55(4):1278–89.

82. Kopp F, Lindenberger U. Effects of joint attention on long-term memory in 9-month-old infants: an event-related potentials study. Dev Sci. 2011;14(4):660–72.

83. Striano T, Reid VM. Social cognition in the first year. Trends Cogn Sci. 2006;10(10):471–6.

84. Hirotani M, Stets M, Striano T, Friederici AD. Joint attention helps infants learn new words: event-related potential evidence. Neuroreport. 2009;20(6):600–5.

85. Wimmer H, Perner J. Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition. 1983;13(1):103–28.

86. Baron-Cohen S, Leslie AM, Frith U. Does the autistic child have a "theory of mind"? Cognition. 1985;21(1):37–46.

87. Auer DP. Spontaneous low-frequency blood oxygenation level-dependent fluctuations and functional connectivity analysis of the "resting" brain. Magn Reson Imaging. 2008;26(7):1055–64.

88. Grosse Wiesmann C, Schreiber J, Singer T, Steinbeis N, Friederici AD. White matter maturation is associated with the emergence of theory of mind in early childhood. Nat Commun. 2017;8:14692.

89. Premack D, Woodruff G. Does the chimpanzee have a theory of mind. Behav Brain Sci. 1978;1:515–26.

90. Perner J, Roessler J. From infants' to children's appreciation of belief. Trends Cogn Sci. 2012;16:519–25.

91. Baillargeon R, Scott RM, He Z. False-belief understanding in infants. Trends Cogn Sci. 2010;14(3):110–8.

92. Choi YJ, Luo Y. 13-Month-olds' understanding of social interactions. Psychol Sci. 2015;26(3):274–83.

93. Kovács ÁM, Téglás E, Endress AD. The social sense: susceptibility to others' beliefs in human infants and adults. Science. 2010;330(6012):1830–4.

94. Southgate V, Senju a CG. Action anticipation through attribution of false belief by 2-year-olds. Psychol Sci. 2007;18(7):587–92.

95. SAJ B, Bernstein DM. What can children tell us about hindsight bias: a fundamental constraint on perspective–taking? Soc Cogn. 2007;25(1):98–113.

96. Bloom P, German TP. Two reasons to abandon the false belief task as a test of theory of mind. Cognition. 2000;77:25–31.

97. Surian L, Caldi S, Sperber D. Attribution of beliefs by 13-month-old infants. Psychol Sci. 2007;18(7):580–6.

98. Moll H, Meltzoff AN. How does it look? Level 2 perspective-taking at 36 months of age. Child Dev. 2011;82(2):661–73.

99. Aichhorn M, Perner J, Kronbichler M, Staffen W, Ladurner G. Do visual perspective tasks need theory of mind? NeuroImage. 2006;30(3):1059–68.

100. Moll H, Tomasello M. Level 1 perspective-taking at 24 months of age. Br J Dev Psychol. 2006;24(3):603–13.

101. Hamilton AF de C, Brindley R, Frith U. Visual perspective taking impairment in children with autistic spectrum disorder. Cognition. 2009;113(1):37–44.

102. Moll H, Kadipasaoglu D. The primacy of social over visual perspective-taking. Front Hum Neurosci. 2013;7(September):558.

103. Schaafsma SM, Pfaff DW, Spunt RP, Adolphs R. Deconstructing and reconstructing theory of mind. Trends Cogn Sci. 2015;19(2):65–72.

104. Dunbar RIM, Shultz S. Evolution in the social brain. Science. 2007;317(5843):1344–7.

105. Fehr E, Fischbancher U. Third-party punishment and social norms. Evol Hum Behav. 2004;25(2):63–87.

106. Camerer CF, Fehr E. When does "economic man" dominate social behavior? Science. 2006;311(5757):47–52.

107. Krueger F, Grafman J, McCabe K. Neural correlates of economic game playing. Philos Trans R Soc Lond Ser B Biol Sci. 2008;363(1511):3859–74.

108. Lee D. Game theory and neural basis of social decision making. Nat Neurosci. 2008;11(4):404–9.

109. Johnson ND, Mislin AA. Trust games: a meta-analysis. J Econ Psychol. 2011;32(5):865–89.

110. Camerer CF, Loewenstein G, Prelec D. Neuroeconomics: How neuroscience can inform economics. J Econ Lit. 2005;43(1):9–64.

111. Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci. 2006;7(4):268–77.

112. McCabe K, Houser D, Ryan L, Smith V, Trouard T. A functional imaging study of cooperation in two-person reciprocal exchange. Proc Natl Acad Sci U S A. 2001;98:11832–5.

113. Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD. The neural correlates of theory of mind within interpersonal interactions. NeuroImage. 2004;22(4):1694–703.

114. Delgado MR, Frank RH, Phelps EA. Perceptions of moral character modulate the neural systems of reward during the trust game. Nat Neurosci. 2005;8:1611–8.

115. King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR, Montague PR. Getting to know you: reputation and trust in a two-person economic exchange. Science. 2005;308:78–83.

116. Delgado MR, Li J, Schiller D, E a P. The role of the striatum in aversive learning and aversive prediction errors. Philos Trans R Soc Lond Ser B Biol Sci. 2008;363(1511):3787–800.

117. Baumgartner T, Heinrichs M, Vonlanthen A, Fischbacher U, Fehr E. Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron. 2008;58(4):639–50.

118. Zak PJ, Kurzban R, Ahmadi S, Swerdloff RS, Park J, Efremidze L, et al. Testosterone administration decreases generosity in the ultimatum game. PLoS One. 2009;4(12):e8330.

119. Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E. Oxytocin increases trust in humans. Nature. 2005;435(June):673–6.

120. Aspé-sánchez M, Moreno M, Rivera MI, Rossi A. Oxytocin and vasopressin receptor gene polymorphisms: role in social and psychiatric traits. Front Neurosci. 2016;9(January):510.

121. van den Bos W, Güroğlu B, van den Bulk BG, Rombouts SA, Crone E. Better than expected or as bad as you thought? The neurocognitive development of probabilistic feedback processing. Front Hum Neurosci. 2009;3(December):52.

122. Mitchell JP. Activity in right temporo-parietal junction is not selective for theory-of-mind. Cereb Cortex. 2008;18(2):262–71.

123. Billeke P, Boardman S, Doraiswamy PM. Social cognition in major depressive disorder: a new paradigm? Transl Neurosci. 2013;4(4):437–47.

124. Billeke P. The more I get to know you, the more I distrust you? Non-linear relationship between social skills and social behavior. Front Psychiatry. 2016;7:49.

125. de Vignemont F, Singer T. The empathic brain: how, when and why? Trends Cogn Sci. 2006;10:435–41.

126. Shenhav A, Botvinick MM, Cohen JD. The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron. 2013;79(2):217–40.

127. Ebitz RB, Platt ML, Ebitz RB, Platt ML. Neuronal activity in primate dorsal anterior cingulate cortex signals task conflict and predicts adjustments in pupil-linked arousal. Neuron. 2015;85(3):628–40.

128. Billeke P, Zamorano F, López T, Rodriguez C, Cosmelli D, Aboitiz F. Someone has to give in: theta oscillations correlate with adaptive behavior in social bargaining. Soc Cogn Affect Neurosci. 2014;9(12):2041–8.

129. Billeke P, Zamorano F, Cosmelli D, Aboitiz F. Oscillatory brain activity correlates with risk perception and predicts social decisions. Cereb Cortex. 2013;23(12):2872–83.

130. Ibáñez MI, Sabater-Grande G, Barreda-Tarrazona I, Mezquita L, López-Ovejero

S, Villa H, et al. Take the money and run: psychopathic behavior in the trust game. Front Psychol. 2016;7(November):1–15.

131. Chang LJ, Smith A, Dufwenberg M, Sanfey AG. Triangulating the neural, psychological, and economic bases of guilt aversion. Neuron. 2011;70(3):560–72.

132. Yoshimura S, Okamoto Y, Onoda K, Matsunaga M, Ueda K, Suzuki S, et al. Rostral anterior cingulate cortex activity mediates the relationship between the depressive symptoms and the medial prefrontal cortex activity. J Affect Disord. 2010;122(1–2):76–85.

133. Damasio AR, Grabowski TJ, Bechara A, Damasio H, Ponto LL, Parvizi J, et al. Subcortical and cortical brain activity during the feeling of self-generated emotions. Nat Neurosci. 2000;3:1049–56.

134. Singer T, Seymour B, O'Doherty J, Kaube H, Dolan RJ, Frith CD. Empathy for pain involves the affective but not sensory components of pain. Science. 2004;303:1157–62.

135. Rilling JK, Sanfey AG. The neuroscience of social decision-making. Annu Rev Psychol. 2011;62:23–48.

136. Camerer CF. Behavioural studies of strategic thinking in games. Trends Cogn Sci. 2003;7:225–31.

137. Cherry T, Frykblom P, Shogren J. Hardnose the Dictator. Am Econ Rev. 2002;92(4):1218–22.

138. Moll J, Krueger F, Zahn R, Pardini M, de Oliveira-Souza R, Grafman J. Human fronto-mesolimbic networks guide decisions about charitable donation. Proc Natl Acad Sci U S A. 2006;103(42):15623–8.

139. Wu S-W, Delgado MR, Maloney LT. The neural correlates of subjective utility of monetary outcome and probability weight in economic and in motor decision under risk. J Neurosci. 2011;31(24):8822–31.

140. Hoffman E, McCabe K, Shachat K, Smith V. Preferences, property rights, and anonymity in bargaining games. Games Econ Behav. 1994;7:346–80.

141. Hutcherson CA, Bushong B, Rangel A. A neurocomputational model of altruistic choice and its implications. Neuron. 2015;87(2):451–62.

142. Raposo A, Vicens L, Clithero JA, Dobbins IG, Huettel SA. Contributions of frontopolar cortex to judgments about self, others and relations. Soc Cogn Affect Neurosci. 2011;6(3):260–9.

143. Kable JW, Glimcher PW. The neurobiology of decision: consensus and controversy. Neuron. 2009;63(6):733–45.

144. Ullsperger M, Fischer AG, Nigbur R, Endrass T. Neural mechanisms and temporal dynamics of performance monitoring. Trends Cogn Sci. 2014;18(5):259–67.

145. Rilling J, Gutman D, Zeh T, Pagnoni G, Berns G, Kilts C. A neural basis for social cooperation. Neuron. 2002;35:395–405.

146. McClure EB, Parrish JM, Nelson EE, Easter J, Thorne JF, Rilling JK, et al. Responses to conflict and cooperation in adolescents with anxiety and mood disorders. J Abnorm Child Psychol. 2007;35(4):567–77.

147. Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD. The neural basis of economic decision-making in the ultimatum game. Science. 2003;300(5626):1755–8.

148. Spitzer M, Fischbacher U, Herrnberger B, Grön G, Fehr E. The neural signature of social norm compliance. Neuron. 2007;56(1):185–96.

149. Apps MAJ, Rushworth MFS, Chang SWC. The anterior cingulate gyrus and social cognition: tracking the motivation of others. Neuron. 2016;90(4):692–707.

150. Shenhav A, Straccia MA, Botvinick MM, Cohen JD. Dorsal anterior cingulate and ventromedial prefrontal cortex have inverse roles in both foraging and economic choice. Cogn Affect Behav Neurosci. 2016;16(6):1127–39.

151. Wittmann MK, Kolling N, Akaishi R, Chau BKH, Brown JW, Nelissen N, et al. Predictive decision making driven by multiple time-linked reward representations in the anterior cingulate cortex. Nat Commun. 2016;7:12327.

152. Kolling N, Wittmann MK, Behrens TEJ, Boorman ED, Mars RB, Rushworth MFS. Value, search, persistence and model updating in anterior cingulate cortex. Nat Neurosci. 2016;19(10):1280–5.

153. Wittmann MK, Kolling N, Faber NS, Scholl J, Nelissen N, MFS R. Self-other mergence in the frontal cortex during cooperation and competition. Neuron. 2016;91(2):482–93.

154. Ruff CC, Fehr E. The neurobiology of rewards and values in social decision making. Nat Rev Neurosci. 2014;15(8):549–62.

155. Apps MAJ, Lesage E, Ramnani N. Vicarious reinforcement learning signals when instructing others. J Neurosci. 2015;35(7):2904–13.

156. Abu-Akel A, Shamay-Tsoory S. Neuroanatomical and neurochemical bases of theory of mind. Neuropsychologia. 2011;49(11):2971–84.

157. Schurz M, Radua J, Aichhorn M, Richlan F, Perner J. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. Neurosci Biobehav Rev. 2014;42:9–34.

158. Saxe R, Xiao D-K, Kovacs G, Perrett DI, Kanwisher N. A region of right posterior superior temporal sulcus responds to observed intentional actions. Neuropsychologia. 2004;42(11):1435–46.

159. Billeke P, Zamorano F, Chavez M, Cosmelli D, Aboitiz F. Functional network dynamics in alpha band correlate with social bargaining. PLoS One. 2014;9(10):e109829.

160. Billeke P, Armijo A, Castillo D, López T, Zamorano F, Cosmelli D, et al. Paradoxical expectation: oscillatory brain activity reveals social interaction impairment in schizophrenia. Biol Psychiatry. 2015;78(6):421–31.

161. Corbetta M, Patel G, Shulman GL. The reorienting system of the human brain: from environment to theory of mind. Neuron. 2008;58(3):306–24.

162. Margulies DS, Ghosh SS, Goulas A, Falkiewicz M, Huntenburg JM, Langs G, et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc Natl Acad Sci. 2016;113:12574–9.

163. Carter RM, Bowling DL, Reeck C, S a H. A distinct role of the temporal-parietal junction in predicting socially guided decisions. Science. 2012;337(6090):109–11.

164. Charness G, Rabin M. Understanding social preferences with simple tests. Q J Econ. 2002;117(3):817–69.