UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y
DOCTORADO EN PSICOLOGÍA

Correlatos Neurales y Verbales del Engaño

T E S I S

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN PSICOLOGÍA

PRESENTA:

Said Enrique Jiménez Pacheco

Tutor Principal:
Dr. Juan José Sánchez Sosa, Facultad de Psicología
Tutores:
Dr. Roberto E. Mercadillo Caballero, Universidad Autónoma Metropolitana
Dr. Eduardo A. Garza Villarreal, Instituto de Neurobiología, UNAM
Dr. Jairo Muñoz Delgado, Instituto Nacional de Psiquiatría, INPRF
Dr. Christian Cruz Torres, Universidad de Guanajuato

CIUDAD DE MÉXICO, AGOSTO 2021

Universidad Nacional Autónoma de México
UNAM
Biblioteca Central
Dirección General de Bibliotecas de la UNAM

UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Índice general

Resumen **3**

Abstract **5**

1 Introducción **7**

2 Neurociencias cognitivas y engaño **11**
2.1 Mecanismos cerebrales del engaño 16
2.2 Procesos emocionales asociados al engaño 19

3 Experimentos conductuales del engaño **23**
3.1 Las palabras son baratas 24
3.2 El experimento del dado 28
3.3 Números aleatorios 30
3.4 Experimentos con promesas 31

4 Psicología de la salud y engaño **33**
4.1 Mecanismos cerebrales de la mentira patológica 36
4.2 Antecedentes 37
4.3 Objetivo 41

5 Métodos **43**
5.1 Participantes 43
5.2 Procedimiento 44
5.3 Diseño 46
5.4 Instrumentos 49
5.5 Adquisición de la fMRI 50
5.6 Análisis Estadístico 50
5.7 Análisis de fMRI 52

6 Resultados **53**
6.1 Resultados de Conducta 53

6.2 Resultados fMRI 58

7 Discusión **63**
7.1 Conclusión 76
7.2 Glosario 81

Referencias **87**

Resumen

Decidir si engañar o no involucra una valoración rápida de consecuencias potenciales y complejas. Además de considerar el beneficio inmediato que se puede obtener mediante del engaño o los resultados negativos de realizarlo, frecuentemente se debe tomar en cuenta la naturaleza de la relación interpersonal de con quien se está interactuando. Si bien la decisión de engañar puede ser recompensante a corto plazo, las consecuencias afectivas y sociales podrían ser devastadoras dependiendo de la cercanía social con la persona afectada. A la fecha, no se ha explorado el efecto que tiene la cercanía social en el engaño y, dado que nuestras interacciones ocurren con personas con quienes compartimos diferentes niveles de cercanía social, es fundamental entender de qué manera esta información puede impactar nuestra conducta. Para dicho propósito, se diseñó un juego económico en el cual se midió el engaño mediante la decisión de romper una promesa y se investigó si esta decisión era modulada por el grado de cercanía social entre la persona que emitió y la que recibió la promesa. Los participantes en el experimento realizaron el juego en un escáner de Imagen por Resonancia Magnética con tres compañeros: 1. Una computadora; 2. Un extraño; 3. Un amigo cercano. Así, se registró la conducta emitida simultáneamente a la función cerebral que le subyace. Los resultados indicaron que la cercanía social alta disminuye la probabilidad de romper una promesa, además, la disminución en la actividad de la corteza anterior del cíngulo y del núcleo estriado ventral se relacionó con la decisión de no cumplir la promesa. Los hallazgos sugieren que la valoración sobre los beneficios de romper una promesa está influida por la familiaridad de la persona afectada, y que dicha valoración involucra la actividad de regiones cerebrales cuya función cognitiva se asocia a la planeación, el procesamiento de recompensas y el aprendizaje social.

Abstract

Deciding whether or not to cheat involves a quick assessment of potential and complex consequences. In addition to considering the immediate benefit that can be obtained through deception or the negative results of doing it, the nature of the interpersonal relationship of the person with whom you are interacting must often be taken into account. Although the decision to cheat can be rewarding in the short term, the affective and social consequences could be devastating depending on the social closeness with the affected person. To date, the effect that social closeness has on deception has not been explored and, given that our interactions occur with people with whom we share different levels of social closeness, it is essential to understand how this information can impact our behavior. For this purpose, an economic game was designed in which deception was measured through the decision to break a promise and it was investigated whether this decision was modulated by the degree of social closeness between the promisor and the promisee. Participants in the experiment played the game on a Magnetic Resonance Imaging scanner with three partners: 1. A computer; 2. A stranger; 3. A close friend. Thus, the behavior emitted simultaneously with the underlying brain function was recorded. The results indicated that high social closeness decreases the probability of breaking a promise, in addition, the decrease in the activity of the anterior cingulate cortex and the ventral striatum was related to the decision not to keep the promise. The findings suggest that the assessment of the benefits of breaking a promise is influenced by the familiarity of the affected person and that such assessment involves the activity of brain regions whose cognitive function is associated with planning, reward processing, and social learning.

Capítulo 1

Introducción

Desde mentir respecto a la apariencia física de nuestra pareja, hasta manipular la información fiscal de una empresa, el engaño es ubicuo en la sociedad y las relaciones interpersonales. Aunque existen muchas maneras de engañar a otras personas, como exagerar la verdad para que luzca como falsa, o simplemente ocultarla (Hippel & Trivers, 2011), quien engaña busca inducir una creencia falsa en los demás influido por diferentes motivaciones (Vrij et al., 2010), ya sea lograr fines de naturaleza extrínseca como obtener una recompensa o evitar un castigo, o satisfacer complejas motivaciones intrínsecas, por ejemplo, mantener un auto-concepto favorable (Mazar et al., 2008).

Independientemente de la forma o la motivación del engaño, el alcance generalizado de este fenómeno ha llamado la atención de científicos de diversas disciplinas empíricas, por lo que en las últimas dos décadas se ha generado un conjunto considerable de hallazgos por medio de diferentes metodologías (Gerlach et al., 2019). Uno de los métodos más frecuentemente utilizados es el autorreporte, tales como encuestas o cuestionarios dirigidos a explorar la ocurrencia de una

manifestación particular del engaño: la mentira. Aunque ésta se considera un tipo o forma en la que el engaño puede acontecer (Dynel, 2016), la literatura neurocientífica y psicológica frecuentemente investiga estos fenómenos de manera no diferenciada. Así, la exploración del engaño con instrumentos de autorreporte, ha sido una de las fuentes más numerosas de información, probablemente por su posibilidad de implementarse en muestras de gran de tamaño.

Por ejemplo, en una encuesta de 1000 adultos estadounidenses, se describió que las personas decían entre una y dos mentiras al día, sin embargo, cuando se les preguntó directamente si mentían, la mayoría afirmó no haberlo hecho en las últimas 24 horas (Serota et al., 2010). Como se podría especular, la distancia entre la cantidad de mentiras que las personas refieren y las que realmente realizan puede ser muy grande, por motivos psicológicos bien documentados como la deseabilidad social (Mazar et al., 2008), o simplemente por la dificultad de reportar con precisión la propia conducta. De modo que, es razonable argumentar que los hallazgos derivados de las metodologías de autorreporte, pueden tener serias limitaciones y que el uso de otro tipo de herramientas de investigación podrían dar lugar a explicaciones más precisas sobre este fenómeno.

En años recientes la economía del comportamiento, las neurociencias y la psicología han realizado intentos para evaluar e investigar el engaño con aproximaciones más objetivas (Gerlach et al., 2019). Algunos de los aportes más notables de la economía del comportamiento han sido las tareas experimentales y los modelos teóricos para describir el engaño, las primeras evalúan este fenómeno con juegos interactivos en los que las personas tienen incentivos para comportarse de manera deshonesta, y los segundos analizan el proceso de toma decisiones que subyace a este tipo de conducta. Entre otros aportes, las neurociencias cognitivas han

brindado las herramientas para identificar los circuitos, estructuras y funciones cerebrales, que constituyen una base fisiológica para que ocurra el engaño; mientras que la psicología, ha interactuado con ambas disciplinas para subrayar el papel, que tienen las dimensiones emocionales y sociales, en los experimentos y modelos que intentan entender el complicado fenómeno del engaño.

Por su carácter generalizado, la información de naturaleza socio-afectiva, le da contexto a las decisiones que pueden impactar en las personas con las que se interactúa. Por ejemplo, si un individuo engaña a otro con el que comparte una relación cercana, las consecuencias podrían ser devastadoras para el vínculo social, quien engaña experimentaría emociones negativas como la culpa o arrepentimiento, además, si el grupo al que pertenecen percibe la falta, probablemente estaría dispuesto a desaprobar y hasta castigar a quien realizó la transgresión. Por otra parte, el comportamiento deshonesto dirigido hacia una persona de poca cercanía social, podría no generar la emoción de culpa en quien engaña, como tampoco tendría el mismo efecto para la persona afectada, ni para los miembros del grupo que atestiguan la violación.

Este fenómeno descrito como el sesgo endogrupal, que se refiere a la preferencia por personas del mismo grupo (con alta cercanía social) en comparación con gente que no pertenece al grupo (sin cercanía social), es un caso que ejemplifica la modulación que puede tener la información de naturaleza social en las decisiones individuales. Aunque se sabe que este sesgo puede influir en decisiones prosociales, como confiar en alguien o responder a la confianza de manera recíproca, se desconoce el papel que pondría tener en la decisión de engañar a otra persona. El presente trabajo resume el panorama de la investigación actual y analiza experimentalmente el efecto de la cercanía social en la decisión de engañar, así como en la actividad

cerebral relacionada con los complejos procesos cognitivos que subyacen a esta decisión.

Capítulo 2

Neurociencias cognitivas y engaño

La neurociencia cognitiva se refiere al estudio de los sustratos neuronales de los procesos mentales. Por lo que el estudio del engaño desde esta disciplina utiliza frecuentemente técnicas de registro y análisis de imágenes cerebrales para entender las bases neurocognitivas de la (des)honestidad humana. Una motivación temprana para el estudio del engaño a nivel neuronal, fue el intento de desarrollar métodos para la detección de la mentira en el marco de interrogatorios forenses (Farah et al., 2014). En este contexto, utilizando pruebas como el denominado *Guilty Knowledge Test* (que consiste en explorar las respuestas que sujetos sospechosos dan a estímulos relevantes sobre un incidente, a los que sólo una persona que hubiera estado involucrada en el evento podría responder), se buscaba identificar marcadores objetivos del engaño, como la activación de respuestas autónomas a nivel fisiológico, o la ocurrencia de estados como ansiedad o culpa a nivel neurocognitivo; lo anterior, bajo la premisa de que sólo alguien que estuviera engañando tendría esas respuestas a piezas de evidencia involucradas en un suceso.

Con hallazgos derivados de investigaciones que utilizan variantes del *Guilty Knowledge Test*, se ha propuesto que el funcionamiento ejecutivo es clave entre los procesos cognitivos asociados con la decisión de engañar. Se ha señalado que esta decisión es tan solo el comienzo de un proceso, que además se compone de otras operaciones mentales que no necesariamente son *"engañosas"* en su naturaleza. Según Sip et al. (2008), los procesos cognitivos que influyen en el comportamiento de engaño son: 1) manejo de información, 2) manejo de la impresión, 3) manejo del riesgo y 4) manejo de la reputación. El **manejo de la información** es crucial para hacer un seguimiento de las creencias falsas que quien engaña intenta inducir en el interlocutor. En este sentido, el manejo de información permite mantener el engaño, lo que implica que quien lo realiza tiene una representación cognitiva respecto a lo que el interlocutor cree que el perpetrador sabe, siente o pretende lograr. Así, el manejo exitoso de la información durante el engaño requiere de la habilidad conocida como teoría de la mente, que consiste en la capacidad de inferir estados mentales en las otras personas, como sus intenciones, deseos o creencias (Grèzes et al., 2006). Este proceso de pensar respecto a la actividad mental de los demás, requiere la activación de una red de regiones cerebrales que incluyen la parte dorsal y ventromedial de la corteza prefrontal, así como la unión temporo-parietal y el surco temporal posterior-superior (Lieberman, 2007).

En cuanto al **manejo de la impresión**, se refiere a que quien implementa el engaño envía señales verbales y no-verbales para generar la confianza de la víctima, que después puede explotar a su favor. Conductualmente, requiere que el perpetrador incurra en el costo de cooperar inicialmente con su víctima, y cognitivamente, necesita monitorear de manera constante el comportamiento de la persona engañada y ajustar el propio para mantener la impresión de

honestidad. Otro de los procesos involucrados en el engaño es el **manejo del riesgo**, que se refiere a ponderar las ganancias y pérdidas de decidir engañar, quien toma esta decisión debe estar seguro de que obtiene más de lo que pierde por su conducta, al mismo tiempo que considera el riesgo de que no se le crea, así como las consecuencias que de ello deriven. Dependiendo de la magnitud del engaño y el daño que ocasione a otros, las consecuencias podrían ocurrir en términos de la pérdida de la reputación, la confianza, el poder o el dinero (Sip et al., 2008). En cuanto a las bases neuronales de lo anterior, se sabe que la corteza medial prefrontal y la corteza anterior del cíngulo, desempeñan un rol relevante en relacionar los comportamientos con sus consecuencias, así como en la codificación tanto del valor esperado de una acción como en la posibilidad de un error (Rushworth et al., 2004). Finalmente, el **manejo de la reputación** hace referencia al esfuerzo cognitivo que realiza el individuo para justificar a sí mismo la necesidad del engaño, al mismo tiempo que utiliza su razonamiento para convencer al grupo de pertenencia de que el engaño tuvo la intención de alcanzar un bien mayor.

Aunque son numerosos los hallazgos de neuroimagen que se han obtenido con el uso de tareas inspiradas en el *Guilty Knowledge Test*, estos experimentos se han criticado porque comúnmente los investigadores solicitan a los participantes que engañen u oculten información, lo que implica que el engaño es instruido y no surge de manera voluntaria y espontánea. Esta restricción en el comportamiento de los sujetos plantea serias limitaciones respecto a la validez de las tareas, al mismo tiempo que representa una barrera, para que los experimentadores estudien el papel de los procesos motivacionales y de toma de decisiones que subyacen a la decisión de engañar. Para hacer frente a estas limitaciones, otro marco conceptual

que aborda la deshonestidad y sus bases neuronales es el de Jenkins et al. (2016), quienes proponen que el engaño es una propiedad de las señales que los organismos se envían entre ellos durante la comunicación, la cual tienen el propósito de ayudar a quien envía la señal con algún fin económico o evolutivo, como la selección de pareja, la reproducción, la evitación de depredadores o el intercambio de bienes.

Desde la perspectiva mencionada, es posible definir con precisión la naturaleza honesta o falaz de una señal (mensaje o comportamiento), en términos de los intercambios entre quien la envía (*señalador*) y la recibe (*receptor*), así como las consecuencias que ocurren en los involucrados. Una señal se considera honesta o confiable, si expresa de manera fidedigna una relación entre alguna característica de la señal y algún atributo del *señalador* o el ambiente, de modo que el *receptor* se beneficiaría de conocer el mensaje en lugar de ignorarlo. Mientras que el engaño, expresa de manera falsa la relación entre las características de la señal y los atributos del mensajero (o su ambiente), sin embargo, en este caso el *señalador* se beneficia de la imprecisión en la información (Searcy & Nowicki, 2010). Por ejemplo, cuando un individuo realiza una promesa respecto a su comportamiento futuro, quien la recibe podría suponer una relación entre el compromiso expresado (la señal) y la conducta futura de quien promete, que le podría ser potencialmente útil para tomar la decisión de depositar su confianza en esa persona. Si la relación entre la promesa y conducta futura es alta, y quien recibe la promesa se beneficia de haber confiado, la señal se clasificaría como honesta; pero si quien promete decide explotar a su favor la confianza del otro, es decir, el *señalador* rompe la relación entre la propiedad de la señal y el atributo que pretende señalar, se consideraría la señal como engaño.

Esta aproximación al estudio del engaño, plantea que los juegos económicos son

una herramienta fundamental para capturar los antecedentes y consecuencias de ambos tipos de decisiones, lo que podría representar las presiones de adaptación que los señaladores y receptores enfrentan usualmente en su medio o contexto; por lo que su uso es recomendado para el desarrollo de tareas experimentales con validez, así como para identificar los procesos cognitivos subyacentes al comportamiento, que desde este marco se pueden diferenciar entre los procesos que pertenecen al *señalador* o al *receptor* (Jenkins et al., 2016). De parte del *señalador* se consideran cuatro procesos cognitivos, (i) construir una representación del problema, que incluye identificar las características de los involucrados, las acciones y los resultados en consideración, (ii) conformar creencias apropiadas (expectativas) sobre la relación entre las posibles acciones y sus consecuencias. Este segundo punto, requiere la habilidad de teoría de la mente, para construir las creencias respecto a los estados mentales del *receptor* y sus futuras acciones.

El *señalador* también realiza la operación de (iii) asignar valores a sus acciones en consideración. Esta asignación de valores puede ponderar no solamente la obtención de ganancias externas (e.g., en el caso de los juegos económicos la ganancia es monetaria), sino también costos psicológicos, como desear evitar la culpa o la desigualdad. Adicionalmente, (iv) el *señalador* debe decidir entre las acciones, de manera que logre un balance entre motivaciones que pueden ser contrarias como lograr un beneficio económico al mismo tiempo que se es honesto, lo cual es posible por medio de procesos como el control cognitivo. En cuanto al *receptor*, un conjunto similar de procesos cognitivos se llevan a cabo: (i) identificación y evaluación de señales, que potencialmente podrían ser honestas o falaces, (ii) anticipar las consecuencias que resultarían de estas señales, (iii)

asignación de valores a las acciones de respuesta posibles y (iv) seleccionar la acción apropiada de acuerdo a los resultados anticipados.

Como brevemente se esbozó, los procesos descritos que influyen en el engaño interpersonal, tienen sustrato en redes corticales y subcorticales involucradas en procesos cognitivos comunes a diversos tipos de interacciones sociales. También, vale la pena mencionar que raramente una región específica en el cerebro humano se asocia con solamente un proceso cognitivo, de modo que las bases neuronales de la toma de decisiones, la valoración del riesgo, el procesamiento de recompensas, la *teoría de la mente*, la inhibición y el control cognitivo, son una lista de algunos de los procesos mentales que incluyen regiones correlacionadas de manera confiable con el engaño.

2.1 Mecanismos cerebrales del engaño

Para establecer los mecanismos cerebrales del engaño, se necesita identificar con precisión las regiones cerebrales implicadas en los procesos específicos que lo definen, por lo que el desarrollo de las técnicas actuales de neuroimagen como la Tomografía por Emisión de Positrones (PET) o la Imagen por Resonancia Magnética funcional (fMRI) han facilitado la medición de la actividad cerebral con significativa resolución espacial y durante la ejecución de tareas. Dos procesos cognitivos que se han investigado con las técnicas de neuroimagen son la inhibición de las respuestas honestas y la generación de respuestas deshonestas (Nobuhito Abe, 2009).

Por ejemplo, un intento por identificar las estructuras cerebrales involucradas en la generación del engaño fue el de Nobuhito Abe et al. (2014), quienes investigaron

decisiones deshonestas que pudieran ser dañinas (*bad lies*) o benéficas para los sujetos (*white lies*). Durante la sesión de registro cerebral por fMRI, se solicitó a los participantes que leyeran escenarios respecto a eventos que pudieran ocurrir en la vida diaria y se les invitó a decidir si mentirían en el caso de que ellos estuvieran experimentando los eventos. Un ejemplo de la condición de *bad lies* fue que los sujetos leyeron una situación en la que imaginaban que habían roto una puerta de un baño público por equivocación y que habían sido cuestionados respecto a su responsabilidad en el asunto. Mientras que un ejemplo de la condición de *white lies* fue una situación en la que los sujetos imaginaban que habían sido despedidos y posteriormente su madre les preguntaba sobre su situación laboral. En ambos casos, los sujetos realizaban la tarea de responder si serían honestos en el escenario hipotético o si dirían una mentira. Los resultados de imagen mostraron que ambos escenarios, en comparación con una condición de control, involucraban el aumento en la actividad cerebral en la región denominada Corteza dorsolateral Prefrontal, mientras que la comparación entre *bad lies* y *white lies* reveló activación de la unión temporoparietal y la corteza medial prefrontal. Los resultados sugieren que las bases neuronales de las decisiones deshonestas son moduladas por factores contextuales como si la decisión puede ayudar o perjudicar a otros (Nobuhito Abe et al., 2014).

A pesar de las diferencias en las tareas experimentales para evaluar el engaño entre los estudios, la participación de las áreas frontales, asociadas con el funcionamiento ejecutivo, se ha relacionado frecuentemente con las respuestas de inhibir o producir el engaño (Nobuhito Abe, 2009). De hecho, áreas corticales como la corteza corteza prefontral dorsolateral, la corteza ventrolateral, la corteza anterior prefrontal y la corteza anterior del cíngulo incrementan su actividad

durante el engaño. Algunas de las funciones cognitivas relevantes para este proceso, son por ejemplo, la manipulación y el monitoreo de la información dentro de la memoria de trabajo, la selección de respuestas y el control cognitivo, las cuales se han relacionado consistenmente con el funcionamiento de la corteza prefontral dorsolateral (MacDonald et al., 2000; Owen, 1997; Rowe et al., 2000).

El engaño requiere alternar entre tareas cognitivas, por ejemplo, quien miente necesita recordar el hecho real y, simultáneamente, crear al menos una versión alternativa a la verdadera. También necesita suprimir de manera activa respuestas asociadas con recompensas, por ejemplo, quien engaña podría inhibir la respuesta facial de felicidad asociada con el estado emocional positivo de la historia verdadera, o requiere inhibir una respuesta automática, por ejemplo, la persona que fingiera una identidad tendría que frenar el impulso de responder con su nombre real cuando se le preguntara. Los tres escenarios anteriores corresponden con los procesos de task switching, reversal learning e inhibición de la respuesta, los cuales se han asociado con el incremento de la actividad en la corteza prefrontal ventrolateral durante tareas de engaño (Chikazoe et al., 2007; Cools et al., 2002; Dove et al., 2000).

También, se piensa que la corteza prefrontal anterior se asocia con la integración de resultados de dos o más operaciones cognitivas para alcanzar un objetivo más alto (Ramnani & Owen, 2004). Por su parte, la corteza anterior del cíngulo se ha implicado en diferentes procesos, como la detección del conflicto y el procesamiento emocional (Kerns et al., 2004; Phan et al., 2002). Así, en virtud de que el engaño se ha asociado con múltiples procesos cognitivos, es probable que las regiones cerebrales mencionadas interactúen entre ellas para conformar este proceso complejo (N. Abe, 2011; Spence et al., 2008).

2.2 Procesos emocionales asociados al engaño

En términos de los procesos emocionales que caracterizan al engaño, se sabe que las personas frecuentemente tienen dificultad para realizar actos deshonestos, particularmente si se trata de conductas que pueden beneficiar al individuo en detrimento de otros (DePaulo & Kashy, 1998). Cuando los sujetos engañan, se incrementa la activación psicofisiológica, específicamente la rama simpática del sistema nervioso autónomo (SNA) responsable de la respuesta de lucha o huída, y cuando farmacológicamente se bloquea esta activación, por medio de la clorpromazina que es un potente antagonista de receptores dopaminérgicos, las personas son incluso más deshonestas. En un estudio, aquellos que tomaron clorpromazina fueron dos veces más deshonestos en un examen que los que tomaron placebo (Schachter & Latane, 1964).

Las respuestas simpáticas del SNA frecuentemente se han asociado con la experiencia de estrés o miedo, y recientemente se ha demostrado que cuando los individuos se involucran repetidamente en conductas deshonestas, el engaño puede incrementar y la experiencia subjetiva de miedo puede disminuir, a la par que se reduce la activación en la amígdala. La amígdala, que se trata de una estructura cerebral subcortical, que tiene un rol primario en el procesamiento de emociones como el miedo y placer, así como con las respuestas conductuales a estos estados afectivos (Ressler, 2010), ha mostrado que reducciones en su actividad durante una decisión deshonesta (medidas con fMRI), predicen incremento de deshonestidad en el futuro. Por lo que se ha propuesto que el mecanismo denominado adaptación, que es susceptible de monitorearse en la actividad de la amígdala, es el que subyace al fenómeno de “pendiente resbaladiza,” que ocurre cuando pequeños

engaños pueden escalar a transgresiones de mayor magnitud (Garrett et al., 2016). A la fecha existen pocos estudios con fMRI encaminados a dilucidar la naturaleza de los procesos emocionales asociados con el engaño. Sin embargo, la investigación en toma de decisiones ha identificado algunos mecanismos interesantes que potencialmente podrían aplicarse en la comprensión de las decisiones deshonestas. Uno de ellos es la denominada *aversión a la culpa*, la que propone que la anticipación de la emoción de culpa en las personas puede motivar el comportamiento de cooperación. La culpa es una emoción negativa que ocurre a consecuencia de la violación a una regla moral personal o social, que se vuelve particularmente notable cuando se cree que se infligió daño, pérdida o malestar a una persona cercana. Motivo por el cual, tratar de evitar la culpa puede ser una fuerte motivación en el proceso de toma de decisiones. En un estudio, las personas que se comportaban de acuerdo con el mecanismo de *aversión a la culpa* en un juego económico de confianza, presentaron incrementos en la actividad de la ínsula, el área motora suplementaria, la corteza dorsolateral prefrontal y la unión temporal-parietal. En tanto, los participantes que maximizaban sus ganancias a costa de otros, tuvieron mayor actividad en la corteza ventromedial prefrontal, la dorsomedial y el núcleo accumbens (Chang et al., 2011).

En el escenario del juego de confianza, ésta se evalúa por medio de cuánto dinero se está dispuesto a invertir en otro individuo, con el riesgo de que este decida no ser recíproco a la confianza otorgada (Berg et al., 1995). Así, la confianza se puede conceptualizar como la creencia de que otra persona cooperará con nosotros, es decir, que estará dispuesta a pagar un costo por ayudar, independientemente de su beneficio personal. En este sentido, si consideramos que el engaño es una forma de explotar la confianza, es probable pensar que los sujetos con menor

aversión a la culpa sean aquellos que tomen más decisiones deshonestas, asimismo, se podría hipotetizar un patrón de actividad cerebral en la corteza ventromedial prefrontal, la corteza dorsomedial y el núcleo accumbens consistente con una estrategia conductual egoísta.

Otro mecanismo para la comprensión del engaño podría ser el de *aversión a la pérdida*, que se desarrolló en el contexto de la Teoría del Prospecto propuesta por Amos Tversky y Daniel Kahneman (Kahneman & Tversky, 1979, 1984). Ésta postula que el valor que los humanos le otorgan a las pérdidas es mayor al valor que se le otorgan a las ganancias (Kahneman & Tversky, 1984). Por ejemplo, el sufrimiento asociado con perder una suma de dinero es mayor que el placer que se experimenta al ganar la misma suma de dinero (Kahneman & Tversky, 1979).

A la fecha, existen algunos estudios que se han orientado a la investigación de las bases cerebrales de la *aversión a la pérdida* (Canessa et al., 2017; Chib et al., 2012; Tom et al., 2007), en los que la corteza ventromedial prefrontal y el cuerpo estriado han mostrado un patrón de decremento en la actividad asociado con las pérdidas mayor al incremento en la actividad asociado con las ganancias (Tom et al., 2007), congruente con el comportamiento del proceso de valoración que se propone en la teoría del prospecto. De forma similar, la amígdala y la ínsula, responden a un patrón específico de las ganancias y las pérdidas, aumentando o disminuyendo su activación y marcando de manera definida la magnitud de las pérdidas potenciales (Canessa et al., 2017).

Así, se puede hipotetizar que algo que empuja a llevar a cabo un comportamiento engañoso sería evitar una pérdida potencial. Aunque el mecanismo de *aversión a la pérdida* se ha estudiado principalmente en términos de decisiones que tienen consecuencias monetarias, es probable que las pérdidas se codifiquen no sólo en

esa dimensión, sino también en términos interpersonales y/o emocionales. En este contexto, estudiar la relación entre el engaño y otros mecanismos regulatorios del comportamiento es un campo relevante y fértil para investigación en neurociencias y psicología.

Capítulo 3

Experimentos conductuales del engaño

Como algunos autores han apuntado de forma certera, la economía y la psicología comparten interés en una amplia gama de temas de la conducta humana y los abordan basados en métodos de laboratorio similares (Ariely & Norton, 2007). Inclusive, áreas que parecieran distantes como la economía conductual y la psicología de la salud, han sumado esfuerzos para afrontar problemas de relevancia social y salud pública, como el VIH, la vacunación, la abstinencia a las metanfetaminas, por nombrar algunos ejemplos (Hanoch & Finkelstein, 2013).

Por lo tanto, no es de sorpresa que, desde el punto de vista de la economía experimental, los hallazgos relacionados con el engaño sean vastos. La óptica clásica de la economía supone que los humanos somos egoístas y que, por lo tanto, mientras hubiera beneficio por mentir o engañar, no tendríamos problema de hacerlo (Crawford & Sobel, 1982). Sin embargo, la evidencia experimental de la economía conductual reciente, indica que el engaño es un asunto más complejo

del que sugiere la propuesta tradicional (Erat & Gneezy, 2012; Fischbacher & Föllmi-Heusi, 2013; Gneezy, 2005; Gneezy et al., 2013).

A la fecha, existen al menos, tres paradigmas experimentales diseñados explícitamente para estudiar las mentiras y el engaño desde la economía conductual, y otro que evalúa el engaño en el contexto de un experimento de confianza y promesas:

1. El juego de palabras baratas entre *remitente* y *recipiente* (Gneezy, 2005).
2. El experimento de tirar un dado de forma encubierta (Fischbacher & Föllmi-Heusi, 2013).
3. El juego de informar sobre un número aleatorio entre $1 - 6$ (Gneezy et al., 2013).
4. Los juegos de confianza con promesas (Charness & Dufwenberg, 2006; Vanberg, 2008).

3.1 Las palabras son baratas

El primer paradigma experimental se planteó en un estudio muy influyente (Gneezy, 2005), el cuál plantea una interacción anónima entre dos personas con asimetría de información: el *remitente* y el *recipiente*. El *remitente* tiene acceso a una matriz de pagos mientras el *recipiente* la ignora, de modo que la asimetría de información ofrece una ventaja para el *remitente* y proporciona un incentivo para explotar la asimetría a su favor mediante el engaño. El *remitente* es informado de dos opciones (A y B) con cantidades que representan pagos para él y su compañero en ambas. La opción A presenta una distribución determinada de pagos entre los jugadores, mientras que la opción B presenta una distribución distinta. El *remitente* debe

enviar un mensaje al *recipiente* respecto a cuál de las dos opciones le hará ganar más dinero, ya que el *recipiente* no tiene acceso a la matriz de pagos, la única información que recibe es la que le brinda el *remitente*.

Una vez que el *remitente* le ha dado el mensaje al *recipiente* respecto a cuál de las dos opciones le hará ganar más dinero, éste último tiene que tomar una decisión referente a si selecciona la opción de pago A o B, y posteriormente los dos jugadores son pagados en función de la decisión del *recipiente*. El engaño simple se evalúa conductualmente cuando el *remitente* le dice al *recipiente* que le conviene más la opción A cuando en realidad le convenía la B. En la versión original de Gneezy (2005), a 75 participantes se les mostró tres tratamientos o condiciones experimentales en las que se variaban los montos de las opciones A y B. En la primera condición experimental, si el *remitente* mentía él ganaba $1 y el *recipiente* perdía $1; en la segunda condición experimental, si el *remitente* mentía él ganaba $1 y el *recipiente* perdía $10; finalmente, en la tercera condición, si el *remitente* engañaba él ganaba $10 y el *recipiente* perdía $10.

En la condición uno, donde el *remitente* ganaba $1 si engañaba y el *recipiente* perdía $1, el 36% de los participantes (27/75) engañó, es decir, le dijo al *recipiente* que le convenía más una opción que en realidad le perjudicaba. En la condición dos, donde el *remitente* ganaba $1 si engañaba y el *recipiente* perdía $10, el número de participantes que engañó disminuyo al 17% (13/75). Por último, en la condición tres, donde la ganancia de engañar era $10 para el *remitente* y una pérdida de $10 para el *recipiente*, el engaño subió al 52% (39/75).

Como se pudo observar en los porcentajes, y como discute Gneezy (2005), los incentivos o ganancias obtenidas por el engaño parecen afectar el comportamiento de las personas, de modo que hay más engaño si se obtienen más ganancias al

hacerlo (por ejemplo $1 vs $10). Sin embargo, este estudio revela también dos fenómenos asociados que no se explican por la teoría del "homo economicus," es decir, por la noción de que los humanos actuamos de forma egoísta y sin consideración del bienestar de las demás personas (Crawford & Sobel, 1982): El primer fenómeno es que las personas no engañan tanto como se pronosticaría por la noción clásica, lo que indica que los remitentes tuvieron consideración respecto a los pagos de los recipientes, o sea, que no solo vieron por su propio beneficio. El segundo fenómeno se refiere a que existe una aversión a mentir, es decir, que las personas se comportan como si hubiera una desutilidad o un costo asociado a la decisión de engañar en sí misma (Gneezy, 2005; Gneezy et al., 2013).

Más específicamente, en la condición tres, en la cual el remitente tenía la oportunidad de ganar más dinero por mentir ($10), fue en la condición en que hubo más participantes que engañaron (39/75). Sin embargo, solo son el 52% del grupo. £Qué pasó con el 48% restante? Gneezy (2005) sugiere que las personas son sensibles a sus ganancias al realizar una mentira, pero también consideran el tamaño de la pérdida que están ocasionando en la otra persona, de modo que en la condición 2, cuando el remitente gana $1 por mentir, pero el *recipiente* pierde $10, se encontró el número más reducido de engaños (17%, 13/75). Y en la condición 1, en la cual el remitente gana $1 y la pérdida del recipiente es $1, el 64% de las personas se abstuvieron de mentir.

Para obtener datos que apoyaran la idea de que las personas tienen aversión a mentir, Gneezy (2005) realizó experimentalmente una versión del juego del *dictador*, el cuál involucraba, nuevamente, la participación de dos personas: el *dictador* y el receptor. El *dictador* decidía entre dos distribuciones de pago, A o B, para él y su compañero (el *recipiente*). En esta versión del juego, el receptor

no realiza ninguna decisión, solo recibe la consecuencia de lo que el *dictador* haya seleccionado. Las distribuciones de los pagos se asignan de forma equivalente con el juego de engaño, es decir, en la condición uno, el *dictador* ganaba $1 y el *recipiente* perdía $1, en la condición dos, el *dictador* ganaba $1 y el *recipiente* perdía $10 y en la condición tres, el *dictador* ganaba $10 y el *recipiente* perdía $10. De modo que si el *dictador* escogía con mayor frecuencia la opción que le beneficiaba a él en detrimento del *recipiente*, sería considerado como evidencia de que las personas tienen *aversión a mentir*.

En los resultados de esta fase, el patrón anterior se mantuvo, pero se incrementó significativamente, en la condición 1 en la que el *dictador* gana $1 y el *recipiente* pierde $1, del 36% en la prueba de engaño pasó a 66% en el juego del dictador, es decir, el 66% en el papel de dictadores escogieron la opción que les beneficiaba más. En la condición 2, en la cual el dictador ganaba$1 y el *recipiente* ganaba $10, del 17% se pasó al 42% de participantes que escogieron la opción que les beneficiaba más en detrimento del *depositario*. En la condición 3, donde el dictador ganaba $10 y el *recipiente* perdía $10, del 52% en el juego de engaño pasó hasta el 90% en el juego del dictador. La frecuencia con la que los dictadores escogieron la opción que más les convenía a ellos en perjuicio del *recipiente*.

La distancia entre los porcentajes de decisión de la opción más conveniente en el juego de engaño en comparación con la decisión en el juego del dictador, es una cuantificación de la aversión a mentir que mostraron los participantes en dicho experimento. De modo que el autor concluye que pareciera que las personas no son indiferentes al proceso por medio del cual obtienen determinado resultado. Es decir, que entre dos pagos matemáticamente idénticos, no habría la misma preferencia entre ellos si para uno tienes que mentir y para el otro no.

3.2 El experimento del dado

El experimento de tirar un dado de forma privada consistió en que los participantes tenían que lanzar un dado de forma encubierta y sin que el experimentador supiera el verdadero resultado del mismo (Fischbacher & Föllmi-Heusi, 2013). Los participantes registraron el valor que obtuvieron del dado. Las personas tenían incentivos para engañar debido a que se les dijo que se les pagaría una cantidad de 1, 2, 3, 4, o 5 CHF (francos suizos) dependiendo del número correspondiente que obtuvieran de lanzar el dado, de modo que si las personas reportaban haber obtenido un 5 el experimentador les pagaba 5 CHF, solo en el caso de que se obtuviera un 6 el pago correspondiente sería 0 CHF.

Según los autores del experimento descrito (Fischbacher & Föllmi-Heusi, 2013), el paradigma experimental que proponen tiene tres ventajas en comparación con otros (por ejemplo el de Gneezy, 2005), las cuales son que: 1. Debido a que el engaño no se puede detectar al nivel individual, los participantes tienen la libertad de comportarse como decidan sin que haya una consecuencia de su conducta; 2. El experimento es muy sencillo de implementar, no llevaba más de 10 minutos por participante; 3. Se conoce, a priori, la distribución verdadera de los resultados del dado si los participantes fueran completamente honestos.

En este experimento, se clasificaba como mentira cuando los participantes reportaban un número distinto al obtenido por lanzar el dado. El estudio se diseñó de forma tal que no era posible distinguir el engaño en un nivel individual, ya que el experimentador realmente no tenía forma de verificar el resultado. De modo que el escenario era propicio para que las personas fácilmente decidieran reportar el número que mejores resultados económicos les diera, independientemente de lo

que hubieran obtenido al lanzar el dado. Sin embargo, dado que se trataba de un dado normal, era viable conocer la distribución esperada respecto a los resultados del dado en condiciones de total honestidad. Así, es posible calcular los resultados reportados por los participantes que se desviaban de la distribución esperada.

Como el lector pudiera anticipar, la cantidad de veces que los participantes reportaban haber obtenido un 5 y que por lo tanto se les pagaban 5 CHF fue la opción que más frecuentemente se reportó. La distribución esperada, si todas las personas fueran totalmente honestas, sería de 16.7% para cada una de las seis caras del dado (1/6). Sin embargo, los resultados no se distribuyeron de forma homogénea. De 389 participantes, el porcentaje de personas que reportó haber obtenido un 5 fue de 35%, para 4 fue 27.2%, para 3 fue 12.6%, para 2 fue 11.6%, para 1 fue 7.2% y el porcentaje de personas que reportó haber obtenido 6 y que por lo tanto se les pagó 0 CHF fue de 6.4%. Así que, números por debajo del 4 fueron significativamente menos reportados que el porcentaje de 16.7% que se esperaba. En tanto, los números 4 y 5 fueron significativamente más reportados que el 16.7% esperado. La distribución de los resultados reportados parece indicar que las personas tendían a decir que habían obtenido un número mayor al que realmente obtuvieron. Además, que hubiera 6.4% de personas que mencionaron haber obtenido 0, (y suponiendo que las personas no mintieron para perder dinero) los autores mencionan que hubo un porcentaje de gente que sí dijo la verdad.

En este experimento se encontró que la mayoría de los sujetos fueron sólo parcialmente deshonestos, ya que mentían respecto al resultado del dado pero no reportaban la opción que maximizaba sus ganancias. Los autores discuten el comportamiento de sus participantes en términos del deseo de aparentar ser

honesto y al mismo tiempo obtener una buena recompensa monetaria, la decisión de reportar un 4 en vez de un 5 podría ser más creíble y también mantenerse por debajo del radar de detección moral propio y el de los demás. Esta explicación, es congruente con la teoría del mantenimiento del autoconcepto propuesta por Mazar et al. (2008), quienes proponen que las personas pueden realizar pequeños actos deshonestos si estos les traen consecuencias positivas, y que por medio del mecanismo de inatención a los estándares morales, las personas son capaces de considerarse a sí mismas como honestas.

3.3 Números aleatorios

Respecto al tercer paradigma experimental para evaluar el engaño (informar sobre un número aleatorio entre 1 y 6) (Gneezy et al., 2013), los investigadores diseñaron una tarea para evaluar específicamente la aversión de los sujetos a mentir. El juego se realizaba en parejas ("A" y "B"). En un ensayo se asignaba aleatoriamente a cada pareja un número entre 1 y 6 que sólo conocía el participante A y éste le decía al B cuál era el número que se les había asignado. Posteriormente, B decidía si creer o no en lo dicho por A. Los pagos para A se incrementaban dependiendo del número que él reportaba y los pagos para B dependían de si confiaba en A y de si el número que les habían asignado coincidía con el estado real de las cosas.

El experimento evalúa la aversión a mentir, porque si el sujeto A sólo estuviera interesado en su propio pago, tendría que reportar a B que siempre se le asignó 6, pues así obtendría el mayor pago. Los investigadores mencionan que la ventaja de este diseño con respecto a otros métodos, es que el pago de A no depende de la reacción de B, y de igual forma afecta el resultado de éste. Los resultados

revelaron que la mayoría de las ocasiones, A reportaba valores mayores que el número asignado, lo que le hacía ganar más dinero, sin embargo no reportaban a B el valor que les daba la ganancia máxima. En otras palabras, mentían en una pequeña proporción y menos del 18% de lo sujetos mintieron cuando las consecuencias para B eran las peores.

3.4 Experimentos con promesas

Finalmente están los experimentos basados en juegos de confianza con promesas (Baumgartner et al., 2009, 2013; Vanberg, 2008). Existen algunas variantes en la implementación de estas tareas, sin embargo la estructura general del juego es relativamente estándar. Se realiza entre dos jugadores el *inversor* y el *destinatario*, el *inversor* tiene originalmente $2 y el *destinatario* no tiene monto. El *destinatario* envía una promesa al *inversor* respecto a que si decide invertir su presupuesto en él, le retribuirá con la mitad de la ganancia, la cuál es mayor que el monto original para el *inversor*. Así, el *inversor* tiene incentivos para dar su dinero al *destinatario*, sin embargo, éste último puede decidir si cumple su promesa y devuelve la mitad al *inversor* o rompe su promesa y se queda con toda la ganancia. Los resultados muestran que existen al menos dos tipos de jugadores en similar proporción, personas que realizan promesas y que las cumplen, y otros que prometen para generar confianza en el *destinatario* y posteriormente explotan dicha confianza. Este paradigma experimental constituye el utilizado en el presente estudio, de modo que será discutido con mayor detalle en las siguientes secciones.

Capítulo 4

Psicología de la salud y engaño

El engaño y las mentiras también ocurren en el contexto clínico. Frecuentemente, el engaño se utiliza como una forma de protección para uno mismo, de modo que es comprensible que éste ocurra en las interacciones entre profesionales de la salud (médicos, enfermeros, trabajadores sociales y/o psicólogos) y sus pacientes. Un ejemplo de engaño en el contexto clínico es el caso de la diabetes mellitus tipo 2; si los pacientes reportaran de forma precisa sus hábitos alimenticios o su adhesión al tratamiento, probablemente no sería necesaria la implementación de múltiples pruebas para monitorear constantemente el nivel de glucosa, hemoglobina glucosilada y/o el estilo de vida (Russell-Minda et al., 2009).

La realidad es que los pacientes y los clínicos frecuentemente (en muchos casos de forma no intencional) dan información que no es precisa o engañan. Se ha descrito que los pacientes pueden tener creencias erróneas respecto a sus síntomas y por lo tanto dar información falsa de su situación, asimismo, pueden ocultar detalles relacionados con sus comportamientos por vergüenza o por proteger a otras personas. Por su parte, los clínicos podrían mentir a un paciente o su familia

sobre la gravedad de una situación de salud, con el fin de no desesperanzarlos o de evitar las emociones asociadas de dar una mala noticia (Wiley, 1998). Así, el engaño forma parte de múltiples interacciones, y como en los siguientes ejemplos, éste es parte central de varios trastornos psiquiátricos que los psicoterapeutas pueden observar en su práctica cotidiana.

Dentro de los trastornos psiquiátricos en los que se ha identificado la presencia del engaño, el Trastorno Límite de la Personalidad (TLP) podría ser un ejemplo clave. Este trastorno se caracteriza por inestabilidad generalizada en el funcionamiento cognitivo, conductual y (principalmente) emocional de quien lo padece. También, son frecuentes las autolesiones, los intentos suicidas y las conductas impulsivas (American Psychiatric Association, 2013). Como estos pacientes oscilan entre extremos de idealización y devaluación de las personas con las que tienen vínculos sociales y/o emocionales, y despliegan esfuerzos frenéticos de evitar el abandono real o percibido, es frecuente el uso del engaño como una estrategia para manipular y mantener cerca a personas que quisieran alejarse de ellas. También, se han descrito pacientes con TLP que mienten a los profesionales de salud respecto a su toma de medicamentos o respecto a su estado emocional, a fin de evitar o conseguir hospitalizaciones para su beneficio (Wiley, 1998).

En otro ejemplo, el engaño parece ser parte fundamental del Trastorno Antisocial de la Personalidad, ya que se caracteriza por escasa capacidad de empatizar con otras personas y por el desafío de las normas sociales (American Psychiatric Association, 2013). Cuando estos pacientes engañan, tienen poca consideración de las consecuencias que su conducta tendría en otras personas. También pueden experimentar poca culpa y, como tienden a presentar impulsividad, pueden engañar de forma frecuente para obtener gratificación inmediata. Se ha descrito

que los sociopatas más habilidosos pueden mentir sin ser detectados e inclusive experimentar placer por engañar a los demás (Board & Fritzon, 2005; Wiley, 1998).

Los pacientes que abusan de sustancias son otro grupo que comúnmente miente. Suelen hacerlo respecto al uso de drogas, o la cantidad que consumen (por ejemplo, de alcohol) y las cosas que hacen para poder obtener las sustancias. Como frecuentemente las drogas de las que se abusa son ilícitas, las personas que consumen engañan para evitar consecuencias legales o como una forma de minimizar el estigma que pudieran experimentar por el consumo (American Psychiatric Association, 2013).

Un diagnóstico psiquiátrico particularmente problemático es el Trastorno Facticio, que se caracteriza porque los pacientes fingen de forma voluntaria y/o se causan alguna enfermedad o lesión, con el objetivo de asumir el rol de enfermo y recibir atención médica. Esto entraña un problema serio por la inversión de tratamientos y otros recursos médicos que podrían haberse otorgado a un paciente que realmente lo necesitara (Yates & Feldman, 2016).

Probablemente, una de las manifestaciones más serias de este diagnóstico es el Síndrome de Munchausen por proximidad, en el que se induce intencionalmente síntomas o enfermedades en otra persona, quien frecuentemente es un niño o alguien dependiente para obtener atención o servicios. Lo anterior es una forma de abuso con posibles consecuencias letales, porque normalmente quien causa el problema es una persona muy cercana (como un padre o la madre) y quien lo recibe está prácticamente indefenso, como en el caso de un infante preverbal o un adulto mayor (Wiley, 1998).

Si bien existen casos clínicos en los que el engaño está presente, frecuentemente no

se dispone de mayor herramienta que la entrevista clínica para su identificación. Así, la implementación de la neuroimagen funcional podría tener un papel primordial en el reconocimiento de estos trastornos clínicos y de los mecanismos neurobiológicos que subyacen al engaño. La siguiente sección describe algunos estudios encaminados a la detección de estos mecanismos.

4.1 Mecanismos cerebrales de la mentira patológica

Identificar los mecanismos cerebrales de la mentira patológica puede ser fundamental tanto para la comprensión de la neurobiología del engaño "normal," como del relacionado con problemas psiquiátricos (N. Abe, 2011). La mentira patológica se puede definir como falsificación completamente desproporcionada de cualquier fin discernible, puede involucrar dominios generalizados y complejos, que se manifiestan por períodos de años, o incluso de por vida, y su manifestación es destructiva para la calidad de vida de quien las sufre y de sus víctimas (Dike et al., 2005).

En una serie de estudios de neuroimagen estructural por MRI, se comparó el volumen de diferentes estructuras cerebrales en un grupo de sujetos con historia de mentiras patológicas, otro con trastorno antisocial de personalidad y un grupo control. Se encontró que los mentirosos patológicos tuvieron mayor volumen de materia blanca prefrontal en comparación con los otros dos grupos, específicamente en la corteza orbitofrontal, la corteza inferior frontal y la media (Yang et al., 2005, 2007). Aunque estos estudios no permiten identificar si las diferencias estructurales

son producto de la mentira patológica o más bien su causa, sí permiten subrayar el papel de la corteza prefrontal relacionado con el engaño (N. Abe, 2011).

Existen dos interpretaciones del hallazgo estructural mencionado, la primera propone que el exceso de materia blanca en el prefrontal tiene el fin de facilitar el procesamiento de la información durante el engaño, mientras que la segunda, indica que el exceso de sustancia blanca es funcionalmente ineficiente, lo que lleva a desinhibición y resulta en la tendencia a mentir. Aunque ambas interpretaciones pudieran tener soporte empírico, la falta de estudios funcionales complica la conclusión respecto al hallazgo anatómico. De modo que los estudios de neuroimagen funcional son un área pendiente en la comprensión de la neurobiología del engaño.

4.2 Antecedentes

Como se ha descrito, una de las ópticas actuales desde donde se estudia el engaño es la neurociencia cognitiva, facilitada por los avances de técnicas de registro y análisis de neuroimagen funcional. Los principales hallazgos en esta línea indican de forma sistemática la participación de la corteza prefrontal, la ínsula y la corteza anterior del cíngulo, entre otras estructuras, para coordinar los diversos procesos sociales, cognitivos y emocionales que conforman el engaño (N. Abe, 2011; Lisofsky et al., 2014).

De forma paralela con los hallazgos de las neurociencias cognitivas, la economía experimental ha otorgado datos sólidos y tareas conductuales ecológicamente válidas para el estudio y comprensión del engaño en circunstancias interactivas (Gneezy, 2005; Gneezy et al., 2013), estos antecedentes y el actual interés por

combinar modelos basados en Teoría de Juegos y los métodos neurocientíficos (Fehr & Camerer, 2007; Sanfey, 2007), han dado origen a algunos estudios que claramente sintetizan la neurociencia y la economía experimental para el estudio del engaño (Por ejemplo, Baumgartner et al., 2009; Volz et al., 2015).

Las aproximaciones metodológicas mencionadas tienen la ventaja de captar la naturaleza dinámica de la interacción social ya que en efecto suponen interacciones reales y no simuladas. Dentro de sus características, los participantes no son forzados a realizar respuestas determinadas, sino que ocurren de forma voluntaria y, adicionalmente, estos estudios representan una forma cuantificable de las consecuencias que pudieran ocurrir tras realizar diferentes tipos de comportamientos, como engañar. Estas ventajas están ausentes en muchas de las propuestas actuales para el estudio del engaño, y representan probablemente las limitaciones más importantes en la investigación del área (N. Abe, 2011; Lisofsky et al., 2014; Sip et al., 2008).

En la agenda del estudio científico del engaño, uno de los tópicos pendientes es la exploración de los contextos sociales que lo facilitan o lo inhiben (N. Abe, 2011). En este sentido, existen numerosas investigaciones que abordan la influencia que tiene la información social en la toma de decisiones y los correlatos cerebrales asociados (Amodio & Frith, 2006; Bhanji & Delgado, 2014). Algunos de los hallazgos más interesantes, apuntan a que la sola presencia de otra persona puede afectar en cómo se procesa la recompensa en el cerebro. Por ejemplo, recompensas monetarias matemáticamente idénticas, tienen diferentes intensidades de respuesta en el cuerpo estriado dependiendo de la cercanía social de la persona con que se comparten; en un estudio, cuando las recompensas se compartieron con un amigo, hubo mayor activación del cuerpo estriado ventral que cuando se compartieron

con un extraño o computadora; además, la experiencia subjetiva de excitación coincidió con el patrón de actividad de la estructura cerebral mencionada (D. S. Fareri et al., 2012).

Otra estructura cerebral que responde a la cercanía social es la corteza medial prefrontal, la cual se ha identificado que desempeña un papel fundamental en la integración de representaciones tanto de acciones potenciales como de sus resultados (Amodio & Frith, 2006). Por ejemplo, se ha encontrado que cuando se compite con una persona de la propia red social (amigo) en comparación con alguien fuera de la red (extraño), la respuesta de la corteza medial prefrontal es más intensa si se trata de un amigo a si se trata de un desconocido (Dominic S. Fareri & Delgado, 2014). Se cree que la corteza medial prefrontal valora diferencialmente la cercanía social, probablemente considera los resultados potenciales de las interacciones con personas de alta cercanía social como más informativos, que los que ocurren con personas de cercanía baja o nula.

Aunque se ha investigado la influencia de la cercanía social en la toma de decisiones sociales como la cooperación o competencia, se desconoce el rol que podría tener este factor en la decisión de engañar así como sus bases cerebrales. En este contexto surge una de nuestras preguntas de investigación, la cual es si habrá algún efecto de la cercanía social en el engaño, es decir, si manipulamos la presentación de compañeros en un juego económico, incluyendo personas que pertenecen a la red social o de alta cercanía social (amigos) e incluyendo personas que no forman parte de la red social o de poca cercanía social (desconocidos) £las personas se conducirán diferencialmente en términos de la frecuencia y magnitud del engaño en función del compañero de juego? Y desde luego, £Cuáles serán los procesos cognitivos y los correlatos neuronales que darán sustrato a sus decisiones?

Para contestar estas preguntas de investigación se plantea un paradigma experimental basado en un juego económico, llamado el Juego de Confianza con Promesas (Berg et al., 1995; Charness & Dufwenberg, 2006), adaptado para su realización durante la evaluación con Imagen por Resonancia Magnética Funcional (fMRI) (Baumgartner et al., 2009). De forma típica, el juego tiene dos jugadores simultáneos, uno de ellos (*inversor*) decide si ofrecer o no su presupuesto a otro (*depositario*). Si el *inversor* decide ofrecer el presupuesto, el investigador multiplica el valor del presupuesto por una cantidad y posteriormente el *depositario* decide cuánto regresarle al *inversor*. La versión del juego de confianza con promesas supone que antes de que el *inversor* decida ofrecer su presupuesto al *depositario*, éste puede enviarle una promesa sobre cuánto dinero le regresará, con el fin de que el *inversor* tenga más confianza en entregarle su dinero.

El engaño se evalúa desde el punto de vista conductual cuando el *depositario* realiza una promesa sobre la cantidad de dinero que regresará y posteriormente, de forma intencional, no la cumple (Baumgartner et al., 2009). Dentro de las manipulaciones experimentales de nuestro estudio, está la presentación aleatoria de tres compañeros de juego distintos, una persona de alta cercanía social (amigo), de baja cercanía social (desconocido) y una computadora como control no humano, quienes llevarán a cabo el papel de *inversor* mientras que el participante evaluado con fMRI desempeñará el papel de *depositario*, quien tendrá que realizar las promesas y tendrá que decidir en qué magnitud y frecuencia cumple la promesa hecha a sus compañeros de juego regresando o no el dinero prometido.

4.3 Objetivo

En virtud de los antecedentes descritos, la presente investigación tuvo el fin de evaluar la influencia de la cercanía social en el engaño, por medio de un experimento en donde se pudieron obtener mediciones del comportamiento en el contexto de un juego de económico, así como de la actividad cerebral asociada con las decisiones.

Capítulo 5

Métodos

El objetivo del presente estudio fue evaluar el efecto de la cercanía social en la decisión de romper una promesa. También, identificar las regiones cerebrales asociados con el procesamiento de la cercanía social y su relación con las decisiones deshonestas.

5.1 Participantes

Se seleccionaron 45 personas voluntarias (15 hombres), estudiantes de licenciatura, en un rango de edad de 19 a 33 años ($\overline{x} = 23.7$, $s = 3.71$), de estado civil soltero. Todos los participantes reportaron no tener historial médico de tipo neurológico ni psiquiátrico al momento de la evaluación. Los participantes acudieron con un amigo considerado cercano por ellos mismos, quienes cumplían con las siguientes características: era una persona del mismo sexo que los participantes, no existía un vínculo familiar y no se trataba de una persona con una relación de tipo sentimental y/o sexual. Los criterios de inclusión y exclusión que debían mostrar

Tabla 5.1: Criterios de Inclusión y Exclusión

Inclusión	Exclusión
Edades entre 18 - 35 años	Historial psiquiátrico y/o neurológico
Sin historial neurológico y/o psiquiátrico	Acompañante familiar
Saber leer y escribir	Pareja sentimental y/o sexual

los participantes para ser considerados dentro del estudio se indican en la 5.1. Todos los participantes y sus respectivos acompañantes firmaron un consentimiento informado en el cual se indicó que su participación era voluntaria, que no representaba riesgo de ningún tipo para su integridad ni la de sus acompañantes y que podían decidir terminar el experimento en cualquier momento sin que ello tuviera repercusiones de ningún tipo. Los procedimientos experimentales se sometieron a aprobación por parte del comité de ética del Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz (INPRF) en la Ciudad de México, así como del comité de investigación y ética de la Universidad Nacional Autónoma de México.

5.2 Procedimiento

1. Reclutamiento de participantes: Se contactó a todos los participantes haciendo publicidad mediante correos electrónicos a alumnos de la Facultad de Psicología de la Universidad Nacional Autónoma de México. Las personas interesadas escribían un correo electrónico y/o se comunicaban mediante teléfono celular con el investigador principal (SJ).
2. Filtrado de participantes: El investigador se comunicaba por teléfono celular con los interesados y les explicaba brevemente que se trataba de un estudio

sobre toma de decisiones. Posteriormente, les preguntaba sobre algunos de sus antecedentes médicos o psiquiátricos, y finalmente, les indicaba que para formar parte del estudio debían acudir con un acompañante, que fuera un amigo, considerado cercano, de su mismo sexo, no familiar, ni pareja sexual y/o sentimental. Si el participante, no estaba de acuerdo o no cumplía (o no podía cumplir) con algún criterio se le agradecía el interés y se descartaba.

3. Valoración psicométrica: Aquellos que cumplían con las características del estudio, se les indicaba que para continuar su participación debían realizar una serie de valoraciones psicométricas, adaptadas para su realización en línea. Las pruebas eran enviadas por correo electrónico y se obtenían los resultados tan pronto como los participantes llenaran las evaluaciones.
4. Día del experimento: Se daba la bienvenida al participante y a su acompañante. En ese momento se les presentaba al extraño, quien era una persona del equipo de investigación que se hacía pasar por otro participante. Se les explicaba verbalmente la estructura del experimento, diciendo que consistía en la realización de un juego adaptado para computadora (ver diseño experimental). Posteriormente, mediante la computadora, se les explicaba el procedimiento del juego y se les permitía realizar entre 3 y 5 ensayos de prueba para que dominaran la ejecución. Para verificar que ambos habían comprendido la tarea, el investigador realizaba preguntas al participante, a su amigo y al extraño, sobre la estructura del juego y sobre qué podían hacer ante determinadas circunstancias hipotéticas. Finalmente, se indicaba al participante que su amigo y el extraño continuarían con su participación en una sala diferente, de modo que el participante iniciaba el juego pensando que lo estaba realizando con sus compañeros.

5. Información para el amigo: En una sala aparte, se indicaba al amigo el objetivo del estudio y se le mencionaba que su función era solamente saber si su mera presencia podía influir en el comportamiento del participante. Asimismo, se le mencionaba que las respuestas que observaba el participante como si fueran del amigo o el extraño habían sido programadas.
6. En el experimento, se le mostraban al participante diferentes pantallas en las que debía realizar una serie de decisiones (ver diseño experimental). Básicamente realizó una promesa sobre la frecuencia de veces que pagaría una cantidad fija y posteriormente, tuvo la posibilidad de cumplir o no su promesa, pagando dicha cantidad ante diferentes compañeros de juego: el amigo, el desconocido y la computadora.
7. Después del experimento, se realizó una sesión breve de conversación con el participante, se le indicaba que las respuestas de sus compañeros o parejas no eran reales con el fin de que él experimentara todas las posibilidades del juego en la misma magnitud. Se les dio la oportunidad de retirar su participación si consideraban que no estaban de acuerdo con alguna de las manipulaciones experimentales.

5.3 Diseño

Los participantes realizaban una tarea de toma de decisiones, que les era presentada como un juego económico. La tarea involucró la participación de tres compañeros con los que los participantes jugaban de forma independiente, los cuales conformaban tres niveles de cercanía social: cercanía alta (cuando jugaban con el amigo), cercanía baja (cuando era el extraño o desconocido) y cercanía

nula (cuando se presentaba una computadora).

La tarea involucró dos condiciones de juego: con promesa (ensayos en que se realizaban promesas y luego se pagaba o no una cantidad fija) y sin promesa (ensayos en que no había promesa sólo se pagaba o no una cantidad fija). Las condiciones fueron implementadas de manera aleatoria, de modo tal que los participantes no anticiparan la identidad de su compañero para el siguiente ensayo y dicho orden de presentación fue exactamente el mismo para todos los participantes, a fin de garantizar que todos experimentaran la tarea de la forma más similar posible. Lo anterior da lugar a un diseño experimental intrasujetos de (2 x 3):

1. Condición de juego: sin promesa y con promesa
2. Cercanía social: nula, baja y alta.

La tarea consistió en 24 ensayos (8 por cada nivel de cercanía social y 12 por la condición de presencia y ausencia de promesas), en los que los participantes, por medio de una interfaz computarizada, realizaban promesas a sus tres compañeros, las cuales indicaban que *siempre*, *casi siempre*, *algunas veces* o *nunca* pagarían la mitad de sus ganancias. Posteriormente, la historia encubierta en el juego, es que el compañero tendría información de la promesa y con base en ello decidiría invertir una cantidad fija en el participante. Sin embargo, las decisiones de los compañeros fueron programadas de antemano, para que la conducta de los compañeros fuera proporcionalmente igual para todos los participantes. Una vez que el participante realizaba la promesa, observaba la supuesta decisión de sus compañeros y sabía si su compañero había invertido o no, así como la identidad del mismo (amigo, desconocido o computadora). Finalmente, el participante podía decidir pagar la mitad o no pagar la mitad (ver Figuras 5.1 y 5.2).

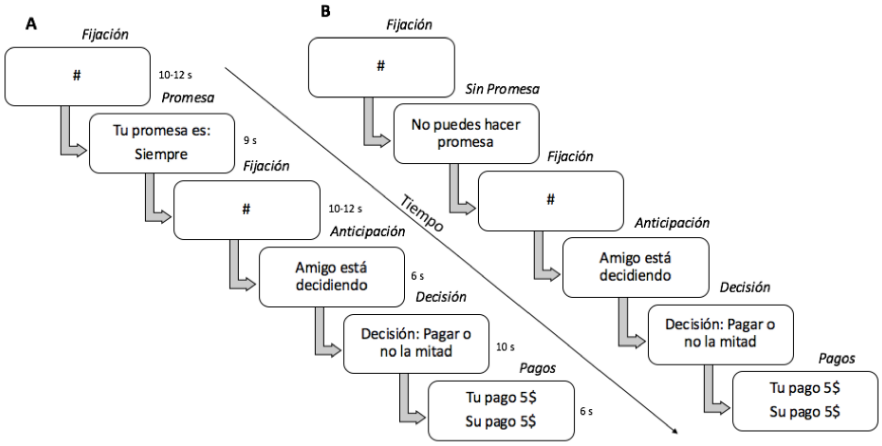

Figura 5.1: Tarea Experimental

Tarea experimental adaptada para su implementación en computadora. En la parte **A** se ve el Juego con promesas en un ensayo típico del juego. En la fase de promesas, el participante escogía entre: *siempre*, *casi siempre*, *algunas veces* o *nunca*, pagaré la mitad. En la fase de anticipación, el participante recibía información sobre con quién de sus compañeros realizaba el ensayo: amigo, desconocido o computadora. En la fase de decisión se le informaba al participante cuál había sido la decisión de su compañero. En caso de que el compañero hubiera invertido, se habilitaba la posibilidad de decidir si se pagaba o no la mitad; en caso de que el compañero no hubiera invertido, se le indicaba tal información y se procedía a la fase de pago. En la fase de pago, se informaban las ganancias que correspondían a cada participante. En la parte **B** se ve el juego sin promesas en un ensayo típico. Todas las fases corresponden, la única diferencia es que durante esta parte los sujetos no pueden realizar promesas (Figura 5.1).

En la Figura 5.2 se muestra la estructura del juego económico. El participante realizaba una promesa, que era informada a su compañero de ensayo, con dicha

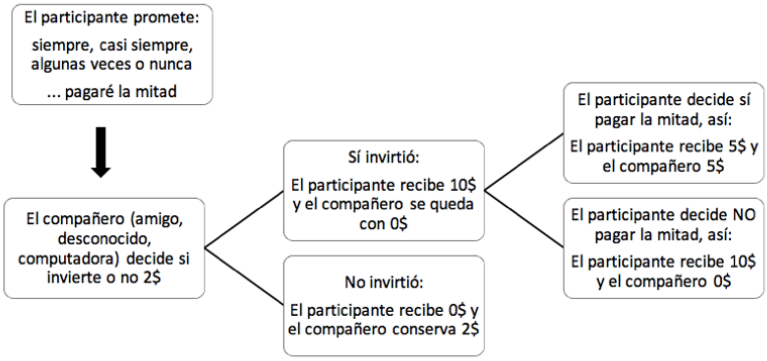

Figura 5.2: Estructura de los pagos en el juego económico

información el compañero decidía si invertir o no su presupuesto de $2. En caso de no invertir, el ensayo culminaba y los pagos eran $2 para el compañero y $0 para el participante, sin embargo, en caso de invertir, la estructura del juego propone multiplicar los $2 por un factor de 5, de modo que el participante recibe $10, finalmente, el participante decide regresar la mitad de su nuevo monto ($5 para el compañero y $5 para sí) o conservar los $10 y dejar a su compañero con $0. La mitad de los ensayos se realizaron sin la fase de promesa, con el objetivo de determinar si la promesa tiene o no un efecto en el comportamiento.

5.4 Instrumentos

El grado de cercanía social que experimenta el participante con relación a sus socios o compañeros de juego se midió utilizando la escala de inclusión del otro en el yo (IOS) (Aron et al., 1992), la cual consiste en siete pares de círculos adyacentes, que están etiquetados con la palabra YO y la palabra OTRO y los cuales varían en función del grado en el que se traslapan indicando cercanía o lejanía percibida con

respecto al otro. La escala de inclusión del otro en el yo IOS, es una forma breve y validada para valorar la cercanía social, así que solicitaríamos a los participantes que realizaran la escala hacia los amigos, los extraños y las computadoras usadas como control no social. Lo mismo se les solicitó a los amigos hacia el participante, el extraño y la computadora, con el fin de verificar que ambos tengan una visión similar de la relación entre ellos y otras personas.

5.5 Adquisición de la fMRI

Las imágenes cerebrales fueron adquiridas en un resonador Phillips Ingenia 3T MR system (Phillips Healthcare, Best, The Netherlands, and Boston, MA, USA) con una bobina de 32 canales dS Head. Los datos funcionales durante la tarea fueron adquiridos utilizando una secuencia de imagen T2*-weigthed echo planar, con tiempo de repetición (TR) / tiempo de eco (TE) = 20000/30 ms, flip-angle = 75ž y recuperación de inversión para la supresión del liquido cerebroespinal. Se adquirieron un total de 510 rebanadas axiales con resolución isotrópica de 3 mm, field of view = 240 mm y matriz de adquisición = 80 x 80. Los datos estructurales, fueron adquiridos por medio de una secuencia T1-weighted (tiempo de repetición = 7 ms, tiempo de eco = 3.5 ms, flip angle = 8ž, field of view = 240 x 240 y voxeles isotrópicos de 1.0 mm).

5.6 Análisis Estadístico

Se obtuvieron las proporciones de las decisiones de pagar o no pagar la mitad, así como las promesas realizadas. Posteriormente, se calcularon las variaciones

en las proporciones de acuerdo con los compañeros de juego y la condición de presencia o ausencia de promesas. Se realizó inferencia Bayesiana para evaluar el efecto de las condiciones experimentales en la tasa de pago a nivel individual y poblacional. Para tal propósito, se realizó un modelo jerárquico que asume que la incertidumbre en el efecto de las promesas y parejas sobre la conducta individual proviene de distribuciones poblacionales para los parámetros del modelo. En este modelo, cada decisión proviene de una distribución Bernoulli con una probabilidad de pagar determinada; el objetivo del modelo jerárquico es predecir cada decisión a través de la combinación lineal de los predictores transformados por su función de enlace inversa *logit*.

El modelo estima que la probabilidad de pagar la mitad para cada ensayo, depende del efecto de la presencia de promesas, así como de los compañeros con cercanía social baja y alta. También asume que pueden existir diferencias individuales en el desempeño durante la tarea, las cuales son asumidas por parámetros específicos que constituyen ajustes a los coeficientes poblacionales.

El modelo se programó en R por medio del paquete `brms` (Bürkner, 2017), que realiza la inferencia mediante el muestreo por Cadenas de Markov Monte Carlo (MCMC) mediante Stan. Las distribuciones posteriores de todos los parámetros fueron aproximadas con cuatro cadenas de 2000 iteraciones cada una: las primeras 1000 iteraciones de cada cadena fueron descartadas (período de calentamiento), para un total de 4000 muestras post-calentamiento. La convergencia del modelo fue evaluada mediante inspección visual de las cadenas y el cálculo del estadístico de diagnóstico de Gelman y Rubin, que para todos los parámetros fue de 1, lo que se interpreta como convergencia.

5.7 Análisis de fMRI

Los datos fueron analizados por medio del software FSL 6.0.1 (FMRIB's Software Library). Las imágenes de cada participante fueron corregidas para movimiento y tiempo, además de normalizadas al espacio MNI (Montreal Neurological Institute, EPI voxel size = 2 X 2 X 2 mm). Los datos fueron suavizados por medio de un filtro Gausiano de 6 mm y filtrados para bandas altas durante el análisis. La señal de fMRI Bold Oxygen Level Dependent BOLD fue examinada durante las fases de la tarea experimental: Promesas, Anticipación y Asignación (ver Figura 5.1), por medio del modelo lineal general (GLM). Para determinar la activación específica de cada fase, además de algunas comparaciones de interés, se evaluaron los contrastes Promesas > 0, Decisión > 0, No promesas > 0, Desconocido > Computadora y Amigo > Computadora. Finalmente, para identificar activaciones a nivel grupal y realizar la inferencia, utilizamos un modelo efectos mixtos con las imágenes obtenidas del primer nivel de análisis de cada sujeto. Se determinó la significancia estadística por medio del criterio de clúster con un umbral de $Z > 2$ y una $p < 0.05$, corregido por mútliples comparaciones por medio de False Discovery Rate (FDR).

Capítulo 6

Resultados

6.1 Resultados de Conducta

6.1.1 Descriptivos generales

Los participantes decidieron pagar la mitad a sus socios en el 69% de los ensayos independientemente de la cercanía social y las promesas, lo que indica una mayoría de decisiones recíprocas en la muestra (Figura 6.1A). Lo anterior es congruente con las promesas realizadas, ya que el 79.6% de estas fueron para los niveles más altos de compromiso: el 41.5% y el 38.1% de las promesas, respectivamente, expresaron que *casi siempre* y *siempre* pagarían la mitad a sus socios. Poco más del 20% de las promesas se distribuyeron entre las opciones que expresan menor compromiso, aunque el 15.3% fue para la opción *a veces* y solo el 5.1% para *nunca* (Figura 6.1B). Llama la atención que la promesa más elegida fuera la opción *casi siempre*, si consideramos que no se prometía a ningún socio en específico, la opción *casi siempre* le daba a los participantes la oportunidad de decidir no ser recíprocos en

algunos ensayos, sin que eso implicara necesariamente falta de honestidad.

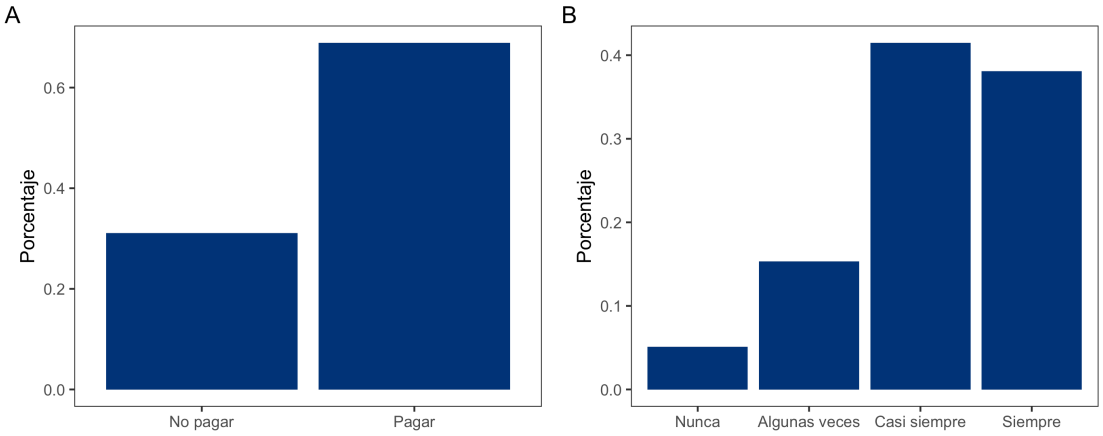

Figura 6.1: Frecuencias de pago y promesas seleccionadas por todos los participantes El panel A) muestra el porcentaje de decisiones recíprocas de todos los participantes, independientemente de la cercanía social y el nivel de promesa. En B) se muestra el porcentaje en que fue seleccionada cada promesa independientemente de los compañeros de juego. La promesa más seleccionada fue casi siempre, que daba la oportunidad de decidir no ser recíproco en algunos ensayos.

6.1.2 Diferencias por cercanía social y promesas

La mayoría de los participantes fueron sensibles a la cercanía social de su compañero, entre más cercanía social mayor fue su proporción de pagos. Además, estas proporciones fueron todavía mayores cuando los sujetos estaban en la condición de promesas en comparación con los ensayos sin promesas. Lo anterior indica la preferencia por el compañero de cercanía alta, a quien le decidieron pagar en el 91% de los ensayos, independientemente de si realizaron o no promesa, también vale la pena mencionar que, sin considerar las promesas, los participantes decidieron ser recíprocos con el extraño en el 67% de los ensayos y en el 48% con la computadora. Los porcentajes anteriores variaron dependiendo de si los

participantes habían realizado una promesa antes, cuando antes se realizó una promesa, el porcentaje de decisiones recíprocas hacia el amigo, subió hasta el 95% de los ensayos, mientras que para el extraño y amigo, los porcentajes de pago fueron de 76% y 53%, respectivamente. En la Figura 6.2 se muestra en azul marino la tasa de pago y en azul claro la diferencia, que corresponde con la proporción de veces que los participantes no pagaron la mitad; se puede observar que la cercanía social incrementa la tasa de pago, así como la condición de promesas.

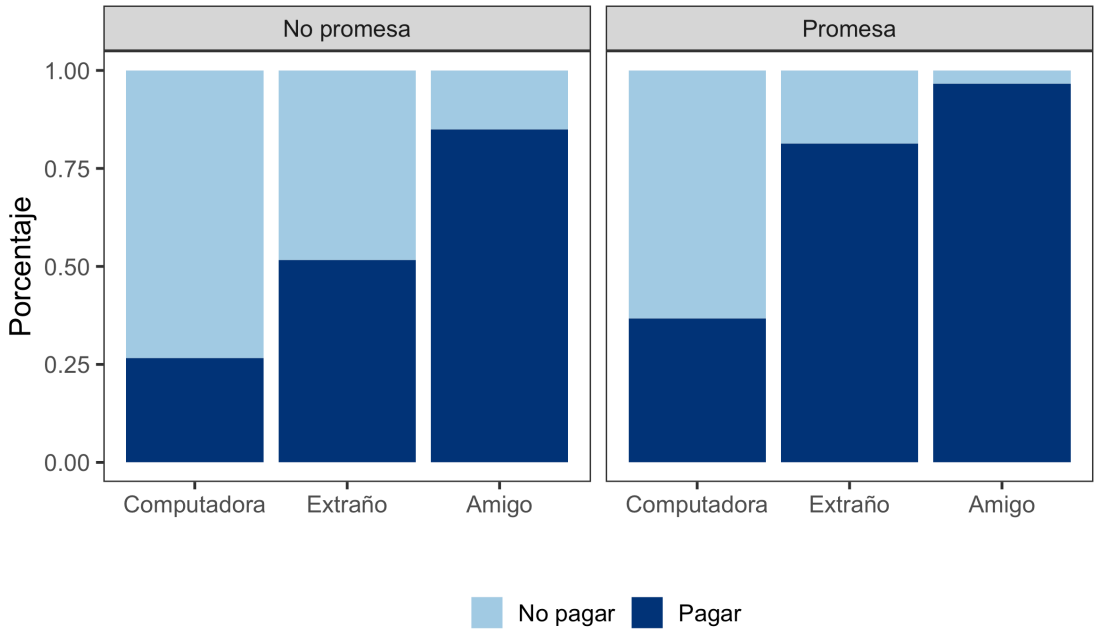

Figura 6.2: Proporciones de pago dependiendo de la cercanía social y la condición de promesas. A la izquierda se muestran las decisiones recíprocas en la condición de no promesas, se puede observar que dependiendo del compañero de juego la proporción de pago fue mayor. En la figura derecha se muestran las decisiones en la condición de promesas, de igual forma, la proporción de pago incrementa en función de la cercanía social del compañero, pero se puede observar que las proporciones fueron mayores en esta condición en contraste con la condición sin promesas.

Tabla 6.1: Estimaciones Bayesianas de los efectos de la cercanía social y promesas en las decisiones recíprocas

Término	Estimación	SE	IC Inferior	IC Superior
Intercepto	-1.62	0.40	-2.31	-1.02
Promesas	1.27	0.40	0.64	1.92
Confederado	2.01	0.55	1.20	2.97
Amigo	4.15	0.77	3.07	5.49

6.1.3 Inferencia Bayesiana

Las condiciones experimentales de promesas y cercanía social (extraño o amigo) incrementaron la probabilidad de pagar la mitad (ver Figura 6.3A). Las pendientes estimadas para los predictores *promesas*, *extraño* y *amigo*, fueron positivas, lo que indica que su presencia incrementa la tasa de decisiones recíprocas. Para interpretar los resultados del modelo empleado, el *intercepto* representa la estimación del efecto conjunto que tuvieron la condición de *no promesas* y la *computadora* sobre la decisión de pagar, por lo que el *intercepto* se puede entender como la condición de referencia, respecto a la cual se determinó la magnitud y dirección de los efectos experimentales.

Prácticamente, todas las condiciones experimentales tuvieron efecto sobre las decisiones con el 95% de probabilidad posterior, sin embargo, el efecto más grande fue el del compañero *amigo*. Además, si se examinan los Intervalos de Credibilidad al 95% (IC) de las estimaciones Bayesianas en la Tabla 6.1, se puede observar que el límite superior del IC del efecto del *extraño* es menor que el límite inferior del IC del efecto del *amigo*, por lo que se puede concluir que existe evidencia robusta de la superioridad del compañero de cercanía social alta, en términos del efecto que tuvo en las decisiones recíprocas de los participantes.

La Tabla 6.1 muestra los efectos poblacionales del modelo Bayesiano jerárquico para la tasa de pago. Se presentan las estimaciones de cada coeficiente del modelo en escala *logit*, su error estándar (SE) y los intervalos de credibilidad (IC) Bayesianos del 95%. Los cuales indican los límites inferiores y superiores de las estimaciones, entre los que se ubica el 95% de probabilidad posterior. Los IC del 95% indican que existe evidencia robusta de los efectos de las condiciones experimentales sobre el comportamiento, ya que todos ellos excluyen el cero como un valor plausible para el coeficiente.

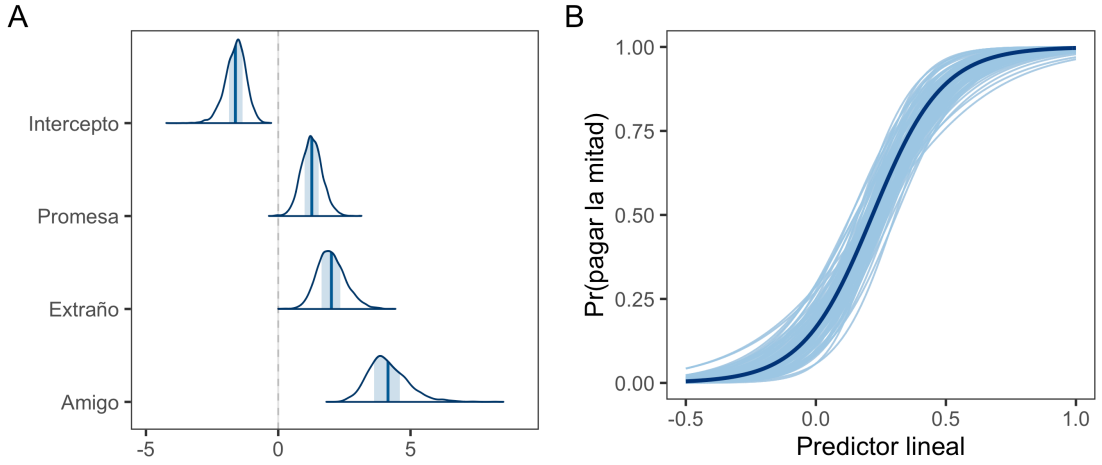

Figura 6.3: Estimaciones Bayesianas de los efectos de las condiciones experimentales

La Figura 6.3A indica las distribuciones de probabilidad posterior de los coeficientes que indican el efecto de las condiciones experimentales: *promesas*, *extraño* y *amigo*. En la abscisa se muestra el tamaño de los coeficientes en escala *logit* y en la ordenada se incluyen los nombres de los predictores evaluados. Se puede observar que todas las distribuciones excluyen confiablemente el cero, lo que indica evidencia fuerte de su efecto sobre el comportamiento de los participantes. Valores mayores al cero indican que la condición experimental aumenta la

probabilidad de regresar el dinero, por lo que se puede observar que la cercanía social alta fue el predictor más importante de las decisiones en la tarea.

La Figura 6.3B muestra la combinación de los efectos de las condiciones experimentales en un mismo predictor lineal y su relación con la decisión de pagar. El predictor lineal incrementa su valor de acuerdo con la cercanía social y la condición de promesas, por lo que se puede observar que a medida que éste incrementa también lo hace la probabilidad de pagar. Se puede notar que cuando el predictor lineal es igual a cero se estima una probabilidad de pagar muy baja, sin embargo a valores a penas superiores al cero la probabilidad de pagar comienza a incrementar rápidamente, lo que representa la sensibilidad de los sujetos a las condiciones experimentales.

6.2 Resultados fMRI

Mediante el modelo de efectos mixtos ejecutado en el cerebro completo se identificó activación significativa en algunas regiones cerebrales para las fases de la tarea: promesa y decisión. Asimismo, se detectó la activación durante la anticipación de los compañeros: computadora, extraño y amigo; posteriormente se exploraron las diferencias en los mapas de actividad cerebral entre el extraño y la computadora, el amigo y la computadora, y el extraño vs. el amigo. Estos contrastes fueron diseñados para explorar la activación cerebral relacionada con el momento de la planeación al realizar una promesa, con la inhibición de respuestas conductuales y emocionales, así como con la evaluación subjetiva de la decisión dependiendo de los niveles de cercanía social.

6.2.1 Efecto de promesas

Respecto a las fases de promesa y no promesa, el modelo de efectos mixtos permitió inferir activación en la Corteza Anterior del Cíngulo bilateral, así como una porción del área premotora. El pico más alto de actividad ocurrió en el lado izquierdo de la Corteza Anterior del Cíngulo, específicamente en las coordenadas MNI $x = 8$, $y = 25$, $z = 28$. La Figura 6.4 muestra en los cortes sagital, coronal y axial, los clusters activados con un puntaje $Z > 2$ y una $p < 0.05$ corregida por clúster.

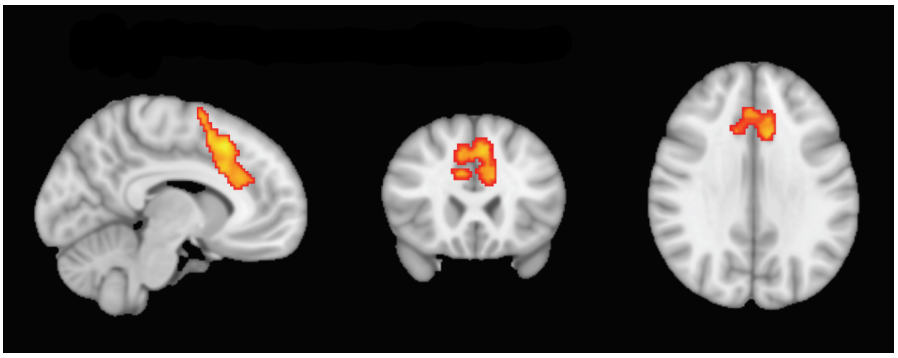

Figura 6.4: Efectos de la condición No promesa en la actividad BOLD. Regiones cerebrales detectadas por el modelo de efectos mixtos durante la fase de promesas en el juego económico. De izquierda a derecha se muestran los cortes cerebrales sagital, coronal y axial en una vista radiológica. El mapa estadístico se superpuso en una imagen estándar MNI de 1 x 1 x 1 mm, el cual muestra en color amarillo brillante las regiones de mayor activación, mientras que el color rojo indica menor activación. El pico máximo de actividad durante la fase de promesas se localizó en la Corteza Anterior del Cíngulo izquierda en las coordenadas x = -8, y = 25, z = 28. Se indican los clusters activados con puntaje Z > 2 y una p < 0.05 corregida por clúster.

6.2.2 Efecto de la decisión de pagar

Respecto a los análisis dirigidos a evaluar la activación durante las decisiones recíprocas, se identificó activación en los ganglios basales izquierdos: el globo pálido

y cuerpo estriado. El pico más alto de actividad se registró en el globo pálido izquierdo, en las coordenadas $x = 16$, $y = 4$, $z = 2$. La Figura 6.5 muestra los conjuntos activados con un puntaje $z > 2$ y una $p < 0.05$ corregida por clúster.

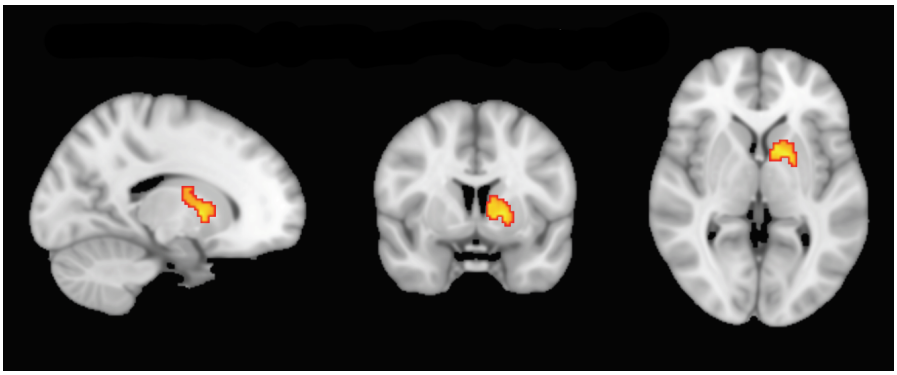

Figura 6.5: Efectos de la decisión de pagar en la actividad BOLD. Regiones cerebrales detectadas por el modelo de efectos mixtos durante las decisiones recíprocas en el juego económico. De izquierda a derecha se muestran los cortes cerebrales sagital, coronal y axial en una vista radiológica. El mapa estadístico se superpuso en una imagen estándar MNI de 1 x 1 x 1 mm, el cual muestra en color amarillo brillante las regiones de mayor activación, mientras que el color rojo indica menor activación. El pico máximo de actividad durante las decisiones se localizó en el globo pálido izquierdo en las coordenadas x = -16, y = 4, z = 2. Se indican los clusters activados con puntaje Z > 2 y una p < 0.05 corregida por clúster.

6.2.3 Efecto de cercanía social durante la anticipación

Con el propósito de evaluar el efecto de la cercanía social durante la anticipación de los compañeros, se contrastaron los mapas de activación del extraño vs. computadora y del extraño vs. amigo. Para las dos comparaciones mencionadas se registró actividad principalmente en el núcleo caudado bilateral, así como en el putamen e ínsula derechos. El pico máximo de actividad se registró en las coordenadas $x = 14$, $y = 12$, $z = 18$ que corresponden con el núcleo caudado

derecho. La Figura 6.6 muestra los clusters activados con un puntaje $z > 2$ y $p < 0.05$ corregida por clúster.

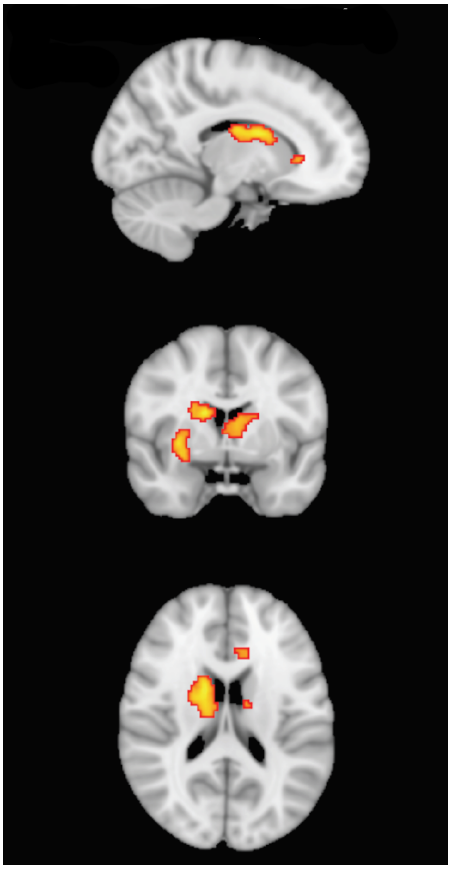

Figura 6.6: Efectos de la cercanía social en la actividad BOLD. Regiones cerebrales detectadas por el modelo de efectos mixtos durante la anticipación de compañeros con cercanía social en comparación con la computadora. De arriba para abajo se muestran los cortes cerebrales sagital, coronal y axial en una vista radiológica. El mapa estadístico se superpuso en una imagen estándar MNI de 1 x 1 x 1 mm, el cual muestra en color amarillo brillante las regiones de mayor activación, mientras que el color rojo indica menor activación. El pico máximo de actividad durante las decisiones se localizó en el núcleo caudado derecho en las coordenadas x = 14, y = 12, z = 18. Se indican los clusters activados con puntaje $Z > 2$ y una $p < 0.05$ corregida por clúster.

Capítulo 7

Discusión

La decisión de engañar, frecuentemente ocurre ante personas que difieren en el grado de cercanía social, como amigos, compañeros de trabajo o vecinos. La información de la familiaridad hacia una persona puede tener un impacto en la frecuencia con la que se actúa conforme a una norma social (por ejemplo, cumplir una promesa a un amigo). Asimismo, la cercanía social puede representarse en estructuras cerebrales que codifican estímulos emocionales y preparan al individuo para emitir respuestas de cooperación o engaño. El presente trabajo tuvo el propósito de explorar los efectos de la cercanía social en el engaño, considerando sus manifestaciones a nivel conductual y en términos de las estructuras cerebrales involucradas.

Dada la complejidad que representa el estudio cuantitativo del engaño y de nuestro problema de interés, en la presente investigación diseñamos una tarea basada en un juego económico con promesas, que a diferencia de otras aproximaciones descritas en la investigación pasada (Baumgartner et al., 2009; Charness & Dufwenberg, 2006; Serota et al., 2010), tuvo la ventaja de poder cuantificar de manera objetiva la

conducta de engaño y no por autorreporte. Permitió que dicha conducta surgiera de manera espontánea por parte de los participantes y no que ocurriera por instrucción del investigador; también, la decisión de engañar se asoció con consecuencias reales para quien lo ejecutó, e involucró la participación de personas significativas para quien realizó la tarea y no simplemente incluyó a sujetos desconocidos.

En nuestro estudio, los participantes decidieron no engañar en la mayor parte de la tarea, lo cual, en términos de la estructura del juego económico, implicó que cumplieron su promesa en la mayoría de los ensayos. Dicho hallazgo coincide con la propuesta de que realizar una promesa puede activar normas sociales de cooperación, que los sujetos tienden a cumplir (Bicchieri, 2002; Fehr & Schurtenberger, 2018). Sin embargo, nuestros datos sugieren que los beneficios o costos de decidir si cumplir dichas normas, o no, se podrían calibrar con respecto a la naturaleza de la relación interpersonal de los involucrados. Aunque las consecuencias monetarias de romper una promesa fueron idénticas para los tres niveles de cercanía social, nuestros participantes prefirieron no romper una promesa cuando la cercanía social fue alta en comparación con cuando fue baja o inexistente.

Así, los datos recopilados indican que la cercanía social es un estímulo que impacta en la decisión de engañar. Específicamente, los sujetos engañaron más del 50% de los ensayos al compañero de cercanía social nula (computadora), engañaron menos del 25% al compañero de cercanía baja (extraño) y menos del 10% al compañero de cercanía social alta. El efecto de la cercanía social alta sobre la decisión de no engañar, fue superior positivo con 95% de probabilidad posterior y fue mayor al efecto que tuvieron los otros compañeros de cercanía social menor. En la literatura de investigación se ha descrito que mentirle a un amigo puede

traer múltiples consecuencias negativas para quien ejecuta y recibe el engaño. Incluso puede haber consecuencias mayores que si el engaño fuera realizado a personas extrañas, por ejemplo, en los grupos se castiga más severamente a sus miembros cuando realizan una transgresión hacia otro individuo del grupo que cuando se realiza hacia una persona que no pertenece al mismo (Linke, 2012). También se ha descrito que romper la promesa a un amigo en comparación con un extraño puede ser devastador emocionalmente para el amigo, puede generar mucha desaprobación del grupo y puede ocasionar culpa, vergüenza y tristeza en quien ha traicionado (Shinada et al., 2004). Así, es plausible que los participantes previeran las posibles consecuencias sociales y/o emocionales de romper las promesas a sus socios, particularmente a sus amigos, lo que en general les permitió tomar decisiones que redujeran la posibilidad de estos resultados indeseables, al mismo tiempo que hicieran más probable la confianza de su socio y que esto se tradujera en consecuencias económicas favorables para los dos.

El comportamiento de los participantes, en cuanto a la preferencia por los individuos de alta cercanía social, nos remite a por lo menos dos de los mecanismos fundamentales de la evolución de la cooperación: la *reciprocidad directa* e *indirecta* (Rand & Nowak, 2013). La *reciprocidad directa* ocurre cuando dos mismos individuos interactúan repetidamente, de modo que las estrategias de comportamiento son condicionales a las consecuencias de interacciones previas, e.g. si cooperaste conmigo ayer entonces yo coopero contigo el día de hoy. La *reciprocidad directa* evoluciona si la probabilidad de nuevas interacciones es suficientemente alta, es decir, si consideramos que el futuro es incierto o ambiguo en términos de las interacciones con los demás, la cooperación de los individuos reduce la incertidumbre al incrementar la probabilidad de que en lo posterior

los demás respondan recíprocamente. En nuestro caso, los individuos cumplieron sus promesas primordialmente hacia sus amigos, de los cuales tenían información disponible respecto a su nivel cooperación y reciprocidad, producto de las múltiples interacciones previas al experimento que probablemente definieron su amistad. Asimismo, aunque durante la tarea los participantes experimentaron la desconfianza de sus amigos en algunos ensayos, la realidad de que las interacciones con ellos trascenderían incluso al contexto del juego económico, pudo representar uno de los factores principales por los que los participantes cumplieron sus promesas en la mayoría de los casos.

Por otra parte, el mecanismo de *reciprocidad indirecta* también podría explicar el comportamiento de los participantes, tanto a los amigos como a los otros dos socios de baja o nula cercanía social. Este mecanismo opera si existen interacciones repetidas dentro de una población, en la que la información respecto al comportamiento de un individuo se esparce para afectar la reputación del mismo, de modo que los demás actúen condicionalmente a esa reputación. En otras palabras, la cooperación ocurre dependiendo de lo que un individuo haga al otro y de lo que los demás digan que hace. Entonces, aunque la cooperación sea costosa, los individuos la realizan para mejorar su reputación e incrementar la posibilidad de recibir ayuda de las demás personas (Rand & Nowak, 2013). En nuestro caso, los individuos cumplieron sus promesas la mayoría de las veces no solo hacia sus amigos, sino que también lo hicieron hacia los extraños; si consideramos que las decisiones de los participantes eran conocidas por todos los socios, es factible que la estrategia de los participantes fuera consolidar una reputación positiva que aumentara la oportunidad de recibir inversiones tanto del amigo como del extraño. Cabe mencionar, que lo anterior no aplicó en el caso de

la computadora, ya que decidieron romper su promesa hacia ella en más del 50% de los ensayos, de modo que se podría inferir que los participantes supusieran que la *reciprocidad directa* o *indirecta*, no operan en el caso del socio no humano.

Si bien, es posible que la *reciprocidad directa* e *indirecta* sean los mecanismos evolutivos que explican el comportamiento de los sujetos durante el experimento, no es factible hacer inferencias respecto a los procesos cognitivo-afectivos que justifican sus decisiones solamente de la observación de su comportamiento. Estudios previos, han sugerido que las personas cumplen sus promesas porque "vencen la tentación" asociada con obtener el beneficio que pudiera ocurrir al incumplir un acuerdo, lo que implicaría que un proceso que se requiere es el control cognitivo, que se usaría para inhibir el impulso de romper la promesa. Otro proceso que podría ser fundamental para mantener una promesa, es la posibilidad de que los sujetos experimenten algún tipo de emoción positiva por comportarse de manera honesta y congruente con sus promesas. De modo que, el uso de imagen por resonancia magnética funcional (fMRI, por sus siglas en inglés) que se empleó durante el estudio, podría ayudarnos a identificar estos procesos y complementar las explicaciones evolutivas.

La imagen cerebral permitió identificar las regiones activadas durante dos momentos fundamentales del juego económico, primero, cuando los sujetos realizaron promesas y posteriormente decidieron cumplirlas; y segundo, el momento en que los sujetos tomaron una decisión hacia alguno de los tres compañeros de diferente cercanía social. Respecto a la fase de promesas, se encontró incremento de la actividad en la corteza anterior de cíngulo, la cual ha sido implicada de manera consistente en el control cognitivo en tareas emocionales (Bhanji & Delgado, 2014), y se ha hipotetizado que su participación durante la

fase de promesas puede señalar la planeación que los sujetos realizan respecto a su comportamiento durante futuras fases de la tarea (Baumgartner et al., 2009). Recientemente se ha subrayado el papel fundamental que la corteza anterior de cíngulo tiene en el aprendizaje social y las normas sociales, se ha propuesto que esta área computa la expectativa respecto al nivel de cooperación de las otras personas (Pärnamets et al., 2020). Considerando lo anterior en su conjunto, el aumento en la actividad de la corteza anterior de cíngulo sugiere que durante la promesa se pondera información respecto a la expectativa del comportamiento potencial de las parejas durante el juego, y se utiliza dicha información para realizar una planeación del nivel de cooperación que el participante exhibirá durante futuras fases.

La corteza anterior del cíngulo es una estructura bilateral que está localizada en la parte anterior y medial de los hemisferios cerebrales, sus neuronas envían axones a las áreas motoras y premotoras, además de poseer conexiones recíprocas con los centros de recompensa en el cerebro, como la corteza orbitofrontal y el cuerpo estriado. La corteza anterior del cíngulo se ha implicado en la atención selectiva a estímulos que, a través del aprendizaje, predicen consecuencias aversivas o recompensantes. Se ha propuesto que esta estructura es uno de los elementos que componen un sistema de valoración para la toma de decisiones sociales y no-sociales, que a través del proceso denominado *error de predicción-recompensa* (reward-prediction error), guía las decisiones hacia la selección de una opción apropiada, la valoración del resultado y el aprendizaje de las consecuencias (Lockwood et al., 2018; Ruff & Fehr, 2014).

El *error de predicción-recompensa* se refiere a la discrepancia entre las consecuencias anticipadas y reales de una decisión determinada, que al

experimentarse ajusta la expectativa del individuo, de modo que la siguiente ocasión en que se tome una decisión particular, la distancia entre la predicción y la consecuencia real sea menor. En el caso de la tarea de promesas, el *error de predicción-recompensa* sugiere que los participantes generan expectativas durante las promesas, por ejemplo, si prometieron que serían recíprocos con sus socios esperan recibir la confianza e inversión de los mismos. La predicción anterior, se actualiza cuando reciben la noticia de que su socio confió o no en ellos; de modo que la actividad de la corteza anterior del cíngulo observada durante la fase de promesa, podría representar la predicción respecto al comportamiento futuro de sus socios.

En cuanto a la actividad cerebral durante la decisión de cumplir la promesa, se encontró aumento de la actividad en una extensa sección del globo pálido izquierdo, particularmente del cuerpo estriado ventral, el cual se ha asociado consistentemente con la representación del valor de una recompensa monetaria (Zhou et al., 2014) o social (D. S. Fareri et al., 2012, 2015; Sip et al., 2015). Específicamente, el cuerpo estriado ventral junto con la corteza ventromedial prefrontal, son dos nodos fundamentales de la red de recompensa, los cuales se han asociado con la anticipación y predicción de las consecuencias estimulantes, así como con el consumo y valoración subjetiva de las mismas (Pujara et al., 2016). Los estudios de neuroimagen funcional han encontrado que la corteza ventromedial prefrontal y el cuerpo estriado exhiben fuerte conectividad funcional durante el reposo, asimismo se activan en conjunto durante tareas de procesamiento de recompensas; y en pacientes con daño bilateral de la corteza ventromedial prefrontal, se ha encontrado que estas lesiones se asocian con disminución de la actividad del cuerpo estriado ventral durante la anticipación de una recompensa,

así como con la disminución del volumen de una subregión de esta estructura denominada núcleo accumbens (Pujara et al., 2016).

Recientemente, se ha propuesto que la corteza ventromedial prefrontal y el núcleo estriado ventral conforman un sistema de valoración común (Pärnamets et al., 2020). Se propone que para tomar cualquier decisión, el cerebro debe asignar y comparar el valor de las opciones disponibles, utilizando un código neuronal independiente de la naturaleza de la recompensa asociada con las opciones, así es posible valorar subjetivamente opciones con consecuencias asociadas de diferentes tipos y utilizar esa información para guiar las decisiones de manera conveniente (Levy & Glimcher, 2012). De acuerdo con lo anterior, la actividad BOLD registrada en el cuerpo estriado ventral durante el presente estudio, sugiere que el valor subjetivo de la opción de cumplir la promesa fue mayor al de no cumplirla, existen estudios que indican que el cuerpo estriado ventral se activa con mayor intensidad cuando las personas actúan de manera justa, cuando cooperan o cuando deciden donar (Cutler & Campbell-Meiklejohn, 2019), de modo que haber actuado conforme a la promesa realizada pudo derivar en mayor utilidad que no haberlo hecho.

De acuerdo con la idea anterior, se podrían proponer varias motivaciones tanto extrínsecas como intrínsecas para explicar la utilidad de proceder de manera congruente con la promesa. Las motivaciones extrínsecas se refieren a consecuencias externas al individuo que estimulan ciertos comportamientos (como la recompensa monetaria); mientras que las motivaciones intrínsecas se pueden inferir cuando un individuo realiza y mantiene una conducta, aunque una recompensa directa no esté presente (Ryan & Deci, 2000). Cumplir la promesa podría generar satisfacción condicionada a comportarse de manera generosa,

o favorecer la propia imagen de que se es una “buena persona,” sin embargo, la misma decisión podría estar influida estratégicamente, por las recompensas futuras que pueden obtenerse vía la reciprocidad o inclusive la evitación de castigo. Nuestra tarea no permite disociar entre las motivaciones intrínsecas o extrínsecas, porque los participantes realizaron varios ensayos con cada socio y no prometieron solamente hacia alguno de ellos en particular, lo que daba la oportunidad de ser estratégico o altruista en diferentes circunstancias. Sin embargo, es muy probable que la decisión de mantener una promesa incluya las dos motivaciones en cierta medida, lo que coincide con la observación reciente de que tanto las decisiones altruistas como estratégicas, activan estructuras cerebrales que se traslapan, como el núcleo accumbens o la corteza anterior del cíngulo (Cutler & Campbell-Meiklejohn, 2019).

Otro argumento que apoya la idea de que tanto la motivación intrínseca como extrínseca pueden explicar la decisión de cumplir la promesa en nuestro estudio, surge si suponemos que los participantes se preocuparon por las creencias que los socios tuvieron respecto a ellos. Cumplir con la promesa puede dar la señal a los demás de que se es un individuo confiable, lo que serviría para dos propósitos por lo menos: 1) favorecería la confianza que los socios pueden tener en el futuro, lo que se traduciría en mayores interacciones de cooperación, asociadas con potenciales recompensas extrínsecas como ganancias económicas y/o sociales; y 2) además de proyectar una imagen confiable, cumplir la promesa podría generar en los demás la creencia de que se es una persona altruista (e.g. que está dispuesta a pagar un precio por dar o ayudar a los otros), lo que según Bénabou & Tirole (2006) puede ser una fuerte motivación intrínseca. Es decir, que los individuos valoran no sólo juzgarse a sí mismos como altruistas, sino que los otros los consideren como

tales. Una predicción basada en lo anterior, sería que si las personas recibieran *más dinero* por cumplir la promesa, probablemente decidirían romperla en más ocasiones. En esas condiciones, introducir un incentivo monetario contingente a una decisión prosocial, podría reducir la credibilidad de que se actúa de manera altruista.

Respecto al efecto de la cercanía social en la actividad cerebral, nuevamente se encontró incremento de la actividad en el cuerpo estriado ventral en el contraste entre computadora y extraño, mientras que el contraste entre extraño y amigo no mostró activación significativa. Se ha propuesto que la activación del cuerpo estriado ventral asociada a la recompensa es sensible a la distinción entre redes sociales, es decir, que entre más cercanía social es mayor la actividad de esta área (D. S. Fareri et al., 2012, 2015; Sip et al., 2015); asimismo, se piensa que compartir una recompensa con alguien de alto grado de cercanía social, podría impactar positivamente en la reafirmación de la relación (Aron et al., 1992; D. S. Fareri et al., 2012). El incremento de la señal BOLD registrada en el cuerpo estriado ventral en el contraste entre extraño y computadora, sugiere que lo sujetos valoran diferencialmente la decisión de cumplir una promesa a un humano en comparación a una máquina, lo que podría entenderse a luz de la utilidad potencial de cumplir una promesa a un compañero con el que puede consolidarse una relación (ej. el extraño), a diferencia de con quien no hay certeza de cooperación próxima (ej. la computadora) (Henrich & Henrich, 2007). Por otro lado, que no se hallara activación significativa en el cuerpo estriado ventral en la comparación entre extraño y amigo fue un resultado inesperado, sin embargo, en el contexto de competencia en contra de otros compañeros con niveles de cercanía social variable, se ha identificado que el cuerpo estriado ventral realiza cómputos

generales asociados con la asignación de valor a las opciones independientemente de la cercanía social, de modo que el resultado de este contraste podría explicarse por el contexto que generó el juego económico, que más que parecer un juego entre socios pudo haberse experimentado como un juego entre competidores (D. S. Fareri et al., 2015).

Una limitación del presente estudio fue el uso de recompensas monetarias hipotéticas en lugar de reales, se podría argumentar que los sujetos podrían no estar suficientemente motivados por las consecuencias de las decisiones en el juego. Sin embargo, en la literatura de investigación sobre toma de decisiones, existen numerosos estudios que han abordado explícitamente la diferencia entre recompensas monetarias hipotéticas versus reales, sin encontrar efectos del tipo de recompensa en tareas de autocontrol, descuento temporal o social (Johnson & Bickel, 2006; Locey et al., 2011). Considerando los estudios mencionados, así como que los hallazgos del presente trabajo tienen congruencia teórica, existen pocas razones para considerar que otro tipo de incentivos hubieran dado lugar a resultados diferentes.

Otra limitación tiene que ver con la falta de anonimato de las decisiones de los sujetos ante el experimentador. Se ha argumentado que la condición anterior podría generar que los participantes tomen decisiones prosociales sólo en apariencia, porque si estuvieran en condiciones de anonimato probablemente no lo harían (Barmettler et al., 2012). Este argumento podría explicar la razón por la que nuestros participantes cumplieron sus promesas en prácticamente todos los ensayos con el amigo, y que también cumplieran en gran medida estos acuerdos con el extraño y la computadora. Si bien es cierto que los investigadores pudieron observar las decisiones de los participantes en todo momento, lo

anterior ocurrió principalmente por el escenario de aplicación del experimento, ya que no es factible realizar la adquisición de imagen cerebral sin monitorear frecuentemente a los participantes. Aunque el experimentador siempre observó a los participantes, la tarea fue implementada directamente con la computadora,por lo que los investigadores nunca manifestaron su presencia durante las decisiones de los sujetos, ni dieron ningún tipo de retroalimentación antes o durante la tarea, que pudiera sesgar de alguna manera las decisiones de los sujetos. Aún así, con los datos recolectados probablemente es imposible diferenciar el efecto de la mera presencia del investigador, y aunque se puede argumentar que fue un contaminante que se mantuvo constante para todos los individuos, detectar el verdadero efecto de la falta de anonimato entre investigador y participante, es una cuestión empírica que debería abordarse en futuros estudios.

Otra crítica que podría hacerse al estudio son los múltiples ensayos con cada uno de los socios. Estrictamente, el juego de confianza usualmente se realiza en una sola exhibición, se argumenta que si los participantes tienen múltiples interacciones con los socios, su comportamiento se puede ver influido por múltiples motivaciones, tanto prosociales como egoístas, de modo que resulta muy complicado, hacer una hipótesis del proceso cognitivo-afectivo y motivacional que subyace a su estrategia. Por ejemplo, si un individuo decide cumplir sus promesas durante los ensayos, no se sabe si lo hace porque tiene una preferencia prosocial incondicional, o bien lo hace para generar una reputación positiva que incremente la confianza de sus socios y que se traduzca en más inversiones para él o ella. La decisión metodológica de realizar múltiples ensayos con cada socio, atiende a la intención de adquirir la señal necesaria para la imagen cerebral funcional; como esta señal es naturalmente débil y muy susceptible de contaminarse, es imposible registrarse con un solo ensayo.

De modo que, aunque se pueden tener dudas respecto a la estrategia cognitiva de nuestros sujetos durante la tarea, por la consistencia de sus decisiones podríamos asumir que se trata de aproximadamente la misma estrategia durante todos sus ensayos. Detectar esta consistencia o *confiabilidad* es necesaria para avanzar en la comprensión de los correlatos cerebrales que dan sustrato a la decisión de mantener o romper una promesa.

El estudio presentado también incluye varias ventajas metodológicas respecto a otras investigaciones sobre el engaño y las promesas. En cuanto al engaño, los estudios encontrados frecuentemente instruyen al participante para que engañe cuando se le solicita, como hemos expuesto anteriormente, esto limita seriamente los componentes cognitivo-afectivos y motivacionales del engaño, por lo que se podría argumentar que la mayoría de los estudios de engaño, en realidad evalúan otros procesos, como la atención, la memoria o inclusive la obediencia. La presente investigación se suma al escaso número de trabajos que utilizan una tarea experimental a la que se le podría adjudicar mayor *validez ecológica*, es decir, que se compone por estímulos y respuestas relevantes en el contexto natural de los individuos, relevantes, en el sentido de que pueden ser funcionales para los propósitos de la supervivencia, la reproducción o la obtención de recursos (Krakauer et al., 2017). Es cierto también, que un juego económico y el escenario limitado de respuestas en un escáner de resonancia magnética podría no considerarse *ecológico* en sí mismo, sin embargo, la posibilidad de que los individuos respondieran libremente y que sus respuestas se relacionaran con consecuencias directas hacia ellos y hacia personas socialmente importantes (como en el caso del amigo), son condiciones metodológicas que destacan en la actualidad entre los paradigmas experimentales para estudiar el engaño y las

decisiones relacionadas con las promesas.

7.1 Conclusión

El presente trabajo resume el panorama actual de la investigación sobre engaño en neurociencias cognitivas, psicología y economía experimental, así como psicología de la salud. Desde neurociencias cognitivas, revisamos que el engaño involucra la participación de la corteza prefrontal, la ínsula y la corteza anterior del cíngulo, entre otras estructuras. Lo que implica que el engaño requiere la coordinación de procesos cognitivos relacionados con toma de decisiones, funciones ejecutivas y procesamiento emocional. Desde el punto de vista experimental se describió que las personas pueden engañar para obtener beneficios económicos y/o personales, sin embargo, los experimentos apuntan de forma reiterada a que, en general, las personas no son insensibles a las consecuencias que el engaño puede tener sobre otros y que en muchas ocasiones, las personas prefieren renunciar a una ganancia que realizar una mentira. También, describimos que el engaño puede ocurrir también en interacciones en el contexto clínico, motivado por proteger el estado emocional de una persona que recibe una mala noticia, para ocultar la poca adhesión a un tratamiento médico, o como parte de diagnósticos psiquiátricos en los que el engaño es parte central, los cuales representan un reto muy complejo para el sistema de salud en general y el abordaje psicoterapéutico en particular.

Desde la definición de engaño y su concepción, sus implicaciones normativas, morales o evolutivas; hasta los procesos cognitivo-afectivos que llevan a los organismos a realizarlo y las implicaciones en la vida social de quienes engañan o son engañados, el estudio y comprensión del engaño es una empresa compleja,

porque las dimensiones de análisis son numerosas y complicadas a su vez. Del presente trabajo derivamos varias lecciones para la investigación del engaño, que pueden ser importantes para la psicología y otros campos relacionados.

Probablemente, una de las lecciones más simples e importantes obtenidas de este trabajo, es la noción de que el engaño es *interactivo*, es decir, que ocurre en el proceso comunicativo entre individuos, y se manifiesta como una característica de las señales o mensajes que se envían de uno a otro, las cuales logran generar creencias falsas en uno que son aprovechadas por el otro. Entonces, un aspecto crucial de la definición de engaño, que no logra ser captado por definiciones tradicionales en psicología, es que el *engañado* paga un precio que el *engañador* cobra, en otras palabras, el engaño ocurre porque tiene consecuencias favorables para quien lo realiza.

Entendido de la forma descrita, el engaño abarca una serie de circuntancias, escenarios o manifestaciones muy diversas y numerosas, desde conductas sencillas como seleccionar fotos que favorecen nuestra imagen y descartar las que nos retratan de manera indeseable, hasta comportamientos serios y criminales, como robar la identidad de una persona o malversar dinero público. En este sentido, prometer algo y después no cumplirlo es totalmente congruente con la definición de engaño, las promesas son señales comunicativas que generalmente funcionan para favorecer la confianza de los demás, si la confianza ocurre y quien prometió se aprovecha de ella, sucede que los demás pagan el precio de haber dado su confianza, y quien prometió resulta beneficiado.

Aunque la investigación de los contratos entre quien promete y recibe la promesa, es tan solo un ejemplo de las situaciones en las que puede ocurrir el engaño, enfocarnos en estudiar esta situación nos otorgó ventajas teóricas y metodológicas

que otras aproximaciones a la fecha no han podido obtener. En la introducción describimos que existe mucha investigación en psicología, economía y neurociencias que hace uso de diversas metodologías como autorreportes, experimentos de comportamiento, o solicitar que los sujetos engañen a conveniencia del investigador mientras se les realiza una valoración por neuroimagen, entre muchas otras. Lo que estas investigaciones tienen en común, es que comprometen la naturaleza interactiva del fenómeno, o si incluyen esta interacción, le restan relevancia a sus hallazgos porque los participantes son anónimos, desconocidos o sin ningún valor afectivo para las personas. O bien, por su metodología estríctamente conductual, la inferencia de los procesos cognitivo-afectivos que subyacen a la decisión de engaño resulta prácticamente imposible. Así, probablemente, el presente se trata de uno de los primeros estudios psicológicos que investiga el engaño en varias de sus dimensiones más complicadas.

A nivel metodológico, se concluye que cualquier investigación experimental del engaño, puede beneficiarse significativamente del uso de juegos económicos, por su capacidad de abstraer la esencia de un sinúmero de interacciones sociales, que ocurren tanto en humanos como en otras especies. Los juegos económicos tienen la ventaja de que permiten analizar objetivamente el comportamiento y motivaciones de quien engaña y es engañado, además de que permiten aislar el efecto de manipulaciones experimentales específicas, como en nuestro caso fue el efecto de la cercanía social en la decisión de engañar.

Respecto a la dimensión conductual, el análisis experimental del comportamiento de engaño, en nuestro caso evaluado por la decisión de romper una promesa, nos indica que las consecuencias monetarias de realizarlo o no, son insuficientes para explicar su ocurrencia. Recompensas matemáticamente idénticas tuvieron

niveles de preferencia muy distintos dependiendo del individuo a quien estuvo dirigido el engaño, de forma tal que la cercanía social se relacionó inversamente con el comportamiento de romper una promesa. Derivado de este hallazgo, resultó necesario analizar diferentes dimensiones que podrían explicar el comportamiento anterior, que van más allá del dominio estríctamente conductual; en el presente documento, se analizaron principalmente algunas de las dimensiones normativo-sociales y cognitivo-afectivas que podrían explicar este fenómeno. Desde el punto de vista normativo-social, es posible dar una explicación del comportamiento observado, en función de normas sociales o reglas, frecuentemente implícitas, que se comparten por los individuos que conforman un grupo, estándares sociales como "no mentir," "cumplir lo que se promete" o "ser recíproco," conforman el marco para decidir rápidamente respecto a un dilema moral como lo es si cumplir o no una promesa. Que nuestros participantes actúen conforme a las normas es esperado y justifica en buena parte sus decisiones en la tarea.

Otro elemento social que explica el comportamiento de nuestros sujetos, es el denomidado *sesgo endogrupal*, que se manifiesta por el trato discriminado entre individuos dentro o fuera del grupo de pertenencia, que puede tomar la forma de preferencia hacia los sujetos del mismo grupo y odio o prejuicio hacia los individuos fuera del grupo. Los participantes cumplieron en gran medida sus promesas, tanto a extraños como a amigos, aunque claramente fue más hacia sus amigos, lo que es congruente con que el *sesgo endogrupal* manifestado en esta tarea fue de preferencia hacia los miembros del grupo, pero no necesariamente de odio hacia las personas fuera del mismo.

En cuanto a la explicación cognitiva-afectiva del comportamiento de nuestros

sujetos, se discutió el papel del *error de predicción-recompensa* que, coherente con la activación de la corteza anterior del cíngulo durante la fase de promesa, sugiere que los individuos generan expectativas respecto al comportamiento de sus socios, que dependiendo de si estas expectativas se cumplen o se frustran, pueden informar la decisión de romper o no la promesa. Lo anterior, es particularmente interesante, en el sentido de que los mecanismos normativos o sociales pueden estar presentes para todos los individuos, pero éstos no se limitan a actuar determinados por los procesos descritos, sino que también aprenden durante la tarea, presumiblemente, a través de la actualización de sus expectativas dependiendo del comportamiento de sus socios, y ajustan sus propias decisiones ponderando la información disponible. No menos importante, es la dimensión afectiva de la decisión de actuar de manera consistente con una promesa, la actividad del cuerpo estríado ventral, nos sugirió que la experiencia de cumplir una promesa a otro humano, puede ser recompensante en sí misma, porque el individuo se puede percibir a sí mismo como alguien justo o altruista, lo que es subjetivamente satisfactorio, al mismo tiempo que señala a otros individuos que se trata de una persona confiable, lo que fortalece su reputación, y cómo se ha descrito, incrementa la probabilidad de que los demás sean cooperadores en el futuro.

Finalmente, es posible decir que el estudio científico del engaño está en etapas iniciales, pero las implicaciones de su comprensión pueden tener un impacto profundo en la vida social, desde su detección en contextos psiquiátricos o criminales, su relevancia para la conformación (o disolución) de vínculos afectivos, hasta su aprobación como una parte innegable de la comunicación. Así, se espera que la presente investigación abone de alguna manera en el entendimiento

psicológico, social, económico y hasta evolutivo, de un fenómeno ubicuo en la naturaleza humana.

7.2 Glosario

Cercanía Social: Se refiere a la distancia física y/o psicológica que un individuo percibe respecto a otro, que puede funcionar para designar la familiaridad hacia esa persona. Una relación social íntima representaría mayor cercanía social, mientras que la intimidad mínima o inexistente reflejaría baja o nula cercanía social. La cercanía social alta incluye el apego y preferencia por la o las personas hacia las cuáles se experimenta, quienes tienen membresía de pertenencia al mismo grupo. Se ha postulado que esta preferencia o *sesgo* hacia personas cercanas socialmente (miembros del mismo grupo) es psicológicamente *primaria*, en el sentido de que se desarrolla antes que las actitudes hacia personas de baja cercanía social (personas fuera del grupo). La presencia de este sesgo, también denominado *sesgo endogrupal*, frecuentemente ocasiona trato diferencial dependiendo de si se pertenece o no al mismo grupo; se propone que esta discriminación podría ocurrir por favoritismo hacia los miembros del mismo grupo, así como prejuicio y hostilidad hacia personas fuera del grupo (Brewer, 2016).

Confianza: Según el modelo diádico de la confianza, ésta es una de las cualidades más importantes para el desarrollo y mantenimiento de relaciones interpersonales, e involucra tres elementos. Primero, la creencia de que las necesidades personales serán satisfechas. Segundo, la inversión de recursos personales, como emociones, tiempo o dinero. Y tercero, la confianza supone un riesgo, ya que la persona en la que se deposita la confianza, podría decidir no cumplir las expectativas o abusar de

la inversión de recursos (Kleinert et al., 2020). Dada la dificultad de *operacionalizar* este constructo, una de las herramientas más utilizadas para medirlo es el juego de confianza, que se trata de un juego económico que representa en su estructura los tres elementos propuestos en el modelo diádico de la confianza.

Cooperación: Una definición simple de cooperación es que un individuo pague un precio para que otro reciba un beneficio. La definición implica que la cooperación ocurre en la interacción entre dos o más individuos, en la cual se invierten recursos de naturaleza tanto intrínseca (e.g. emociones) como extrínseca (e.g. tiempo). Si nos limitamos a la interacción entre dos personas con dos estrategias cada uno (cooperar o desertar), dos cooperadores obtienen mayor beneficio que dos desertores; sin embargo, si uno decide cooperar y el otro no, el individuo que desertó obtiene mayor beneficio por obtener los recursos del otro sin haber invertido los propios, mientras que el cooperador paga el costo más alto. Por lo anterior, para la población es mejor si los individuos cooperan, sin embargo, cada sujeto está tentado a decidir desertar. Una de las herramientas más utilizadas para medir la cooperación es el juego económico conocido como el *dilema del prisionero*, el cual involucra dos jugadores que simultáneamente deben decidir si cooperar o desertar (Rand & Nowak, 2013).

Culpa: se refiere a un estado emocional negativo asociado con la violación de una regla moral o estándar social. La culpa es más notable cuando se cree que se ha infligido daño, pérdida o distrés, a una persona con la que se tiene una relación interpersonal. La culpa cobra relevancia en el contexto de la toma de decisiones, porque se ha propuesto que anticipar la culpa, y querer evitarla, puede ser una fuente de motivación intrínseca muy poderosa (Chang et al., 2011).

Engaño: es una propiedad de las señales que los organismos se envían entre ellos

durante la comunicación, la cual tiene el propósito de ayudar a quien envía la señal con algún fin económico o evolutivo, como la selección de pareja, la reproducción, la evitación de depredadores o el intercambio de bienes. La naturaleza honesta o falaz de una señal (mensaje o comportamiento), se puede definir en términos de los intercambios entre quien la envía (*señalador*) y la recibe (*receptor*), así como las consecuencias que ocurren en los involucrados. Una señal se considera honesta o confiable, si expresa de manera fidedigna una relación entre alguna característica de la señal y algún atributo del *señalador* o el ambiente, de modo que el *receptor* se beneficiaría de conocer el mensaje en lugar de ignorarlo. Mientras que el engaño, expresa de manera falsa la relación entre las características de la señal y los atributos del mensajero (o su ambiente), sin embargo, en este caso el *señalador* se beneficia de la imprecisión en la información (Searcy & Nowicki, 2010).

Juego de confianza: es un juego económico que se utiliza para *operacionalizar* la confianza en términos de los intercambios económicos que realizan dos jugadores. El juego involucra a dos personas con dos estrategias cada una, una de las personas tiene que decidir si confía o no en la otra, y en caso de que la confianza ocurra, la otra persona decide si responder con reciprocidad, o aprovecharse de la confianza otorgada. La confianza se mide a través de la cantidad de dinero que un individuo está dispuesto a invertir en el otro, mientras que la reciprocidad se refiere a cuánto dinero el otro está dispuesto a regresar al inversor. El juego está diseñado de manera tal que el inversor puede acceder a mayores ganancias si confía en el otro, mientras que el beneficiario obtendría la máxima ganancia si decide no ser recíproco con su inversor. Aunque los comportamientos de los dos involucrados son interesantes, en el sentido de que permiten explorar diferentes procesos psicológicos subyacentes,

la decisión del beneficiario está supeditada a la decisión del inversor, ya que si este prefiriera no confiar, el juego terminaría en ese momento (Berg et al., 1995).

Mentira: A menudo los investigadores hacen uso de la palabra engaño o mentira de manera indiscriminada. De acuerdo con la definición de engaño considerada en este trabajo, que lo considera una propiedad de las señales que los organismos se envían durante la comunicación, se puede entender la mentira como una manifestación verbal del engaño, que intenta crear una creencia en el interlocutor, que el comunicador considera falsa. Según esta definición, mentir es un acto intencional y es diferente a otras emisiones incompletas de un mensaje, por ejemplo, cuando un sujeto reporta de manera errónea un suceso porque no lo recuerda con claridad es diferente a mentir (Vrij et al., 2010).

Norma social: Una norma social es una regla de comportamiento en determinada situación, que es compartida, aprobada y conocida por la mayoría de los miembros de un grupo. Generalmente, si las normas sociales se obedecen, generan un ambiente predecible y ordenado, que favorece la cooperación entre personas del mismo grupo. Asimismo, cuando las normas sociales no se siguen, suelen generar desaprobación y castigo por parte de las personas del grupo, incluso cuando no están siendo directamente dañadas por quien transgrede la norma. Es importante subrayar que una norma social no está definida por el comportamiento real de las personas o de sus motivos subyacentes, sino que se define exclusivamente en términos del estándar conductual normativo (Fehr & Schurtenberger, 2018).

Promesa: en economía se dice que las “palabras son baratas” (cheap talk) porque el mero hecho de comunicarse incrementa la cooperación y disminuye el riesgo percibido en un intercambio de bienes entre individuos. Lo anterior ocurre sin necesidad de mayor regulación, sanciones o medidas coercitivas, simplemente la

cooperación acontece porque se prometió. Así, una promesa es la manifestación de un compromiso, el mensaje explícito de que el comportamiento futuro estará determinado por el apego a lo prometido. Se piensa que las promesas hacen salientes las normas sociales, por ejemplo, puede activar la creencia de que "todos van a cumplir con su parte" o bien normas respecto a estándares personales de comportamiento, como "uno debería siempre mantener su palabra." Las personas tienden a cumplir sus promesas porque hacerlo tiene consecuencias emocionales positivas, genera aprobación del grupo y favorece la cooperación en el futuro, al mismo tiempo que no hacerlo, podría generar culpa, castiga y mala reputación (Bicchieri, 2002).

Prosocial: Técnicamente, los comportamientos prosociales pueden ser entendidos como aquellos en los que los individuos incurren en costos para proveer beneficios a los demás, aunque esta es la definición que usamos para cooperación, el comportamiento prosocial es una categoría de mayor jerarquía, que incluye tanto a la cooperación como al castigo altruista. Mientras la cooperación supone incurrir en un costo para ayudar a alguien más, el castigo altruista consiste en que los individuos paguen un costo por castigar a otro individuo, a fin de lograr que su comportamiento se ajuste a una norma. En teoría, ambos comportamientos son prosociales porque benefician al grupo y a su adecuado funcionamiento (Henrich & Henrich, 2007).

Teoría de Juegos: se refiere a la formalización matemática de la conducta estratégica y la interacción social. Una interacción dada se representa por un conjunto de jugadores, las opciones disponibles de cada jugador y el pago que cada jugador obtiene dependiendo tanto de su decisión como la de los otros jugadores. El *dilema del prisionero* y el *juego de confianza* son ejemplos de este tipo de

juegos que describen un problema como el de la cooperación y la confianza.

Traición: es la acción de defraudar o violar un pacto de confianza entre individuos, una transgresión a la norma social de "cumplir con lo que se acuerda o promete." La gravedad de la traición se puede evaluar en términos del daño que ocasiona a quien se traiciona, así como de la intencionalidad que subyace en quien rompe el acuerdo. Realizar traición puede generar consecuencias negativas directas tanto intrínsecas como extrínsecas; por ejemplo, internamente el individuo puede experimentar intensas emociones de culpa o vergüenza, mientras que externamente, podría merecer el castigo de la persona a la que traicionó, e inclusive personas que no fueron dañadas directamente por la traición podrían desaprobar y castigar al transgresor.

Referencias

Abe, Nobuhito. (2009). The neurobiology of deception: evidence from neuroimaging and loss-of-function studies. *Current Opinion in Neurology*, *22*(6), 594–600. https://doi.org/10.1097/WCO.0b013e328332c3cf

Abe, N. (2011). How the Brain Shapes Deception: An Integrated Review of the Literature. *The Neuroscientist*, *17*(5), 560–574. https://doi.org/10.1177/1073858410393359

Abe, Nobuhito, Fujii, T., Ito, A., Ueno, A., Koseki, Y., Hashimoto, R., Hayashi, A., Mugikura, S., Takahashi, S., & Mori, E. (2014). The neural basis of dishonest decisions that serve to harm or help the target. *Brain and Cognition*, *90*, 41–49. https://doi.org/10.1016/j.bandc.2014.06.005

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. https://doi.org/10.1176/appi.books.9780890425596.744053

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci*, *7*(4), 268–277. https://doi.org/10.1038/nrn1884

Ariely, D., & Norton, M. I. (2007). Psychology and experimental economics: A gap in abstraction. *Current Directions in Psychological Science*, *16*(6), 336–339. https://doi.org/10.1111/j.1467-8721.2007.00531.x

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

Barmettler, F., Fehr, E., & Zehnder, C. (2012). Big experimenter is watching you! Anonymity and prosocial behavior in the laboratory. *Games and Economic Behavior*, *75*(1), 17–34. https://doi.org/10.1016/j.geb.2011.09.003

Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., & Fehr, E. (2009). The Neural Circuitry of a Broken Promise. *Neuron*, *64*(5), 756–770. https://doi.org/10.1016/j.neuron.2009.11.017

Baumgartner, T., Gianotti, L. R. R., & Knoch, D. (2013). Who is honest and why: Baseline activation in anterior insula predicts inter-individual differences in deceptive behavior. *Biological Psychology*, *94*(1), 192–197. https://doi.org/10.1016/j.biopsycho.2013.05.018

Berg, J., Dickhaut, J., & McCabe, K. (1995). *Trust, Reciprocity, and Social History* (Vol. 10, pp. 122–142). https://doi.org/10.1006/game.1995.1027

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, *96*(5), 1652–1678. https://doi.org/10.1257/aer.96.5.1652

Bhanji, J. P., & Delgado, M. R. (2014). The social brain and reward: Social information processing in the human striatum. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(1), 61–73. https://doi.org/10.1002/wcs.1266

Bicchieri, C. (2002). Covenants without Swords. *Rationality and Society*, *14*(2), 192–228. https://doi.org/10.1177/1043463102014002003

Board, B. J., & Fritzon, K. (2005). Disordered personalities at work. *Psychology, Crime and Law*, *11*(1), 17–32. https://doi.org/10.1080/10683160310001634304

Brewer, M. B. (2016). Intergroup Discrimination: Ingroup Love or Outgroup Hate?

The Cambridge Handbook of the Psychology of Prejudice, 90–110. https://doi.org/10.1017/9781316161579.005

Bürkner, P.-C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Canessa, N., Crespi, C., Baud-Bovy, G., Dodich, A., Falini, A., Antonellis, G., & Cappa, S. F. (2017). Neural markers of loss aversion in resting-state brain activity. *NeuroImage*, *146*(November 2016), 257–265. https://doi.org/10.1016/j.neuroimage.2016.11.050

Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011). Triangulating the Neural, Psychological, and Economic Bases of Guilt Aversion. *Neuron*, *70*(3), 560–572. https://doi.org/10.1016/j.neuron.2011.02.056

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, *74*(6), 1579–1601. https://doi.org/10.1111/j.1468-0262.2006.00719.x

Chib, V. S., De Martino, B., Shimojo, S., & O'Doherty, J. P. (2012). Neural Mechanisms Underlying Paradoxical Performance for Monetary Incentives Are Driven by Loss Aversion. *Neuron*, *74*(3), 582–594. https://doi.org/10.1016/j.neuron.2012.02.038

Chikazoe, J., Konishi, S., Asari, T., Jimura, K., & Miyashita, Y. (2007). Activation of right inferior frontal gyrus during response inhibition across response modalities. *Journal of Cognitive Neuroscience*, *19*(1), 69–80. https://doi.org/10.1162/jocn.2007.19.1.69

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the Neural Mechanisms of Probabilistic Reversal Learning Using Event-Related Functional Magnetic Resonance Imaging. *Journal of Neuroscience*, *22*(11), 4563–4567.

https://doi.org/10.1523/jneurosci.22-11-04563.2002

Crawford, V. P., & Sobel, J. (1982). Strategic Information Transmission. *Econometrica*, *50*(6), 1431–1451. https://doi.org/10.2307/1913390

Cutler, J., & Campbell-Meiklejohn, D. (2019). A comparative fMRI meta-analysis of altruistic and strategic decisions to give. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2018.09.009

DePaulo, B. M., & Kashy, D. A. (1998). Everyday Lies in Close and Casual Relationships. *Journal of Personality and Social Psychology*, *74*(1), 63–79. https://doi.org/10.1037/0022-3514.74.1.63

Dike, C. C., Baranoski, M., & Griffith, E. E. H. (2005). *Pathological lying revisited*.

Dove, A., Pollmann, S., Schubert, T., Wiggins, C. J., & Yves Von Cramon, D. (2000). Prefrontal cortex activation in task switching: An event-related fMRI study. *Cognitive Brain Research*, *9*(1), 103–109. https://doi.org/10.1016/S0926-6410(99)00029-4

Dynel, M. (2016). Comparing and combining covert and overt untruthfulness. *Pragmatics & CognitionPragmatics and Cognition*, *23*(1), 174–208. https://doi.org/10.1075/pc.23.1.08dyn

Erat, S., & Gneezy, U. (2012). White Lies. *Management Science*, *58*(4), 723–733. https://doi.org/10.1287/mnsc.1110.1449

Farah, M. J., Hutchinson, J. B., Phelps, E. a., & Wagner, A. D. (2014). Functional MRI-based lie detection: scientific and societal challenges. *Nature Publishing Group*, *15*(2), 123–131. https://doi.org/10.1038/nrn3665

Fareri, D. S., Chang, L. J., & Delgado, M. R. (2015). Computational Substrates of Social Value in Interpersonal Collaboration. *Journal of Neuroscience*, *35*(21), 8170–8180. https://doi.org/10.1523/JNEUROSCI.4775-14.2015

Fareri, Dominic S., & Delgado, M. R. (2014). Differential reward responses during competition against in- and out-of-network others. *Social Cognitive and Affective Neuroscience*, *9*(4), 412–420. https://doi.org/10.1093/scan/nst006

Fareri, D. S., Niznikiewicz, M. A., Lee, V. K., & Delgado, M. R. (2012). Social Network Modulation of Reward-Related Signals. *Journal of Neuroscience*, *32*(26), 9045–9052. https://doi.org/10.1523/jneurosci.0610-12.2012

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences*, *11*(10), 419–427. https://doi.org/10.1016/j.tics.2007.09.002

Fehr, E., & Schurtenberger, I. (2018). Normative foundations of human cooperation review-article. *Nature Human Behaviour*, *2*(7), 458–468. https://doi.org/10.1038/s41562-018-0385-5

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise-an experimental study on cheating. *Journal of the European Economic Association*, *11*(3), 525–547. https://doi.org/10.1111/jeea.12014

Garrett, N., Lazzaro, S. C., Ariely, D., & Sharot, T. (2016). The brain adapts to dishonesty. *Nature Neuroscience*, *19*(12), 1727–1732. https://doi.org/10.1038/nn.4426

Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin*, *145*(1), 1–44. https://doi.org/10.1037/bul0000174

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, *95*(1), 384–394. https://doi.org/10.1257/0002828053828662

Gneezy, U., Rockenbach, B., & Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior and Organization*, *93*, 293–300. https://doi.org/

10.1016/j.jebo.2013.03.025

Grèzes, J., Berthoz, S., & Passingham, R. E. (2006). Amygdala activation when one is the target of deceit: Did he lie to you or to someone else? *NeuroImage*, *30*(2), 601–608. https://doi.org/10.1016/j.neuroimage.2005.09.038

Hanoch, Y., & Finkelstein, E. A. (2013). Health psychology meets behavioral economics: Introduction to special issue. *Health Psychology*, *32*(9), 929–931. https://doi.org/10.1037/hea0000009

Henrich, J., & Henrich, N. (2007). *Why Humans Cooperate* (pp. 1–280).

Hippel, W. von, & Trivers, R. (2011). The evolution and psychology of self-deception. *The Behavioral and Brain Sciences*, *34*(1), 1-16; discussion 16-56. https://doi.org/10.1017/S0140525X10001354

Jenkins, A. C., Zhu, L., & Hsu, M. (2016). Cognitive neuroscience of honesty and deception: a signaling framework. *Current Opinion in Behavioral Sciences*, *11*, 130–137. https://doi.org/10.1016/j.cobeha.2016.09.005

Johnson, M. W., & Bickel, W. K. (2006). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the Experimental Analysis of Behavior*, *77*(2), 129–146. https://doi.org/10.1901/jeab.2002.77-129

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292. https://doi.org/10.2307/1914185

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*(4), 341.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior Cingulate Conflict Monitoring and Adjustments in Control. *Science*. https://doi.org/10.1126/science.1089910

Kleinert, T., Schiller, B., Fischbacher, U., Grigutsch, L. A., Koranyi, N., Rothermund, K., & Heinrichs, M. (2020). The Trust Game for Couples (TGC): A new standardized paradigm to assess trust in romantic relationships. *PLoS ONE*, *15*(3), 1–17. https://doi.org/10.1371/journal.pone.0230776

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). *Neuroscience Needs Behavior: Correcting a Reductionist Bias* (No. 3; Vol. 93, pp. 480–490). https://doi.org/10.1016/j.neuron.2016.12.041

Levy, D. J., & Glimcher, P. W. (2012). *The root of all value: A neural common currency for choice*. https://doi.org/10.1016/j.conb.2012.06.001

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology*, *58*, 259–289. https://doi.org/10.1146/annurev.psych.58.110405.085654

Linke, L. H. (2012). Social Closeness and Decision Making: Moral, Attributive and Emotional Reactions to Third Party Transgressions. *Current Psychology*, *31*(3), 291–312. https://doi.org/10.1007/s12144-012-9146-1

Lisofsky, N., Kazzer, P., Heekeren, H. R., & Prehn, K. (2014). Investigating socio-cognitive processes in deception: A quantitative meta-analysis of neuroimaging studies. *Neuropsychologia*, *61*(1), 113–122. https://doi.org/10.1016/j.neuropsychologia.2014.06.001

Locey, M. L., Jones, B. A., & Rachlin, H. (2011). Real and hypothetical rewards in self-control and social discounting. *Judgment and Decision Making*, *6*(6), 552–564. http://journal.sjdm.org/11/11405/jdm11405.pdf

Lockwood, P. L., Wittmann, M. K., Apps, M. A. J., Klein-Flügge, M. C., Crockett, M. J., Humphreys, G. W., & Rushworth, M. F. S. (2018). Neural mechanisms for learning self and other ownership. *Nature Communications*, *9*(1). https:

//doi.org/10.1038/s41467-018-07231-9

MacDonald, A. W., Cohen, J. D., Andrew Stenger, V., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*. https://doi.org/10.1126/science.288.5472.1835

Mazar, N., Amir, O., & Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*, *45*(6), 633–644. https://doi.org/10.1509/jmkr.45.6.633

Owen, A. M. (1997). *The functional organization of working memory processes within human lateral frontal cortex: The contribution of functional neuroimaging*. https://doi.org/10.1111/j.1460-9568.1997.tb01487.x

Pärnamets, P., Shuster, A., Reinero, D. A., & Van Bavel, J. J. (2020). A Value-Based Framework for Understanding Cooperation. *Current Directions in Psychological Science*, 096372142090620. https://doi.org/10.1177/0963721420906200

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). *Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI*. https://doi.org/10.1006/nimg.2002.1087

Pujara, M. S., Philippi, C. L., Motzkin, J. C., Baskaya, M. K., & Koenigs, M. (2016). Ventromedial prefrontal cortex damage is associated with decreased ventral striatum volume and response to reward. *Journal of Neuroscience*, *36*(18), 5047–5054. https://doi.org/10.1523/JNEUROSCI.4236-15.2016

Ramnani, N., & Owen, A. M. (2004). *Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging*. https://doi.org/10.1038/nrn1343

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, *17*(8), 413–425. https://doi.org/10.1016/j.tics.2013.06.003

Ressler, K. J. (2010). Amygdala Activity, Fear, and Anxiety: Modulation by Stress. *Biological Psychiatry*, *67*(12), 1117–1119. https://doi.org/10.1016/j.biopsych.2010.04.027

Rowe, J. B., Toni, I., Josephs, O., Frackowiak, R. S. J., & Passingham, R. E. (2000). The prefrontal cortex: Response selection or maintenance within working memory? *Science*. https://doi.org/10.1126/science.288.5471.1656

Ruff, C. C., & Fehr, E. (2014). *The neurobiology of rewards and values in social decision making* (No. 8; Vol. 15, pp. 549–562). https://doi.org/10.1038/nrn3776

Rushworth, M. F. S., Walton, M. E., Kennerley, S. W., & Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences*, *8*(9), 410–417. https://doi.org/10.1016/j.tics.2004.07.009

Russell-Minda, E., Jutai, J., Ph, D., Psych, C., Speechley, M., Bradley, K., Sc, B. H., Chudyk, A., Sc, M., & Petrella, R. (2009). Health Technologies for Monitoring and Managing Diabetes : *Journal of Diabetes Science and Technology*, *3*(6), 1460–1471.

Ryan, R., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sanfey, A. G. (2007). Social decision-making: insights from game theory and neuroscience. *Science (New York, N.Y.)*, *318*(5850), 598–602. https://doi.org/10.1126/science.1142996

Schachter, S., & Latane, B. (1964). Crime, cognition, and the autonomic nervous system. *Nebraska Symposium on Motivation.*

Searcy, W. A., & Nowicki, S. (2010). *The Evolution of Animal Communication Reliability and Deception in Signaling Systems: Reliability and Deception in*

Signaling Systems (pp. 1–280). Princeton University Press.

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in America: Three studies of self-reported lies. *Human Communication Research*, *36*(1), 2–25. https://doi.org/10.1111/j.1468-2958.2009.01366.x

Shinada, M., Yamagishi, T., & Ohmura, Y. (2004). False friends are worse than bitter enemies: "Altruistic" punishment of in-group members. *Evolution and Human Behavior*, *25*(6), 379–393. https://doi.org/10.1016/j.evolhumbehav.2004.08.001

Sip, K. E., Roepstorff, A., McGregor, W., & Frith, C. D. (2008). Detecting deception: the scope and limits. *Trends in Cognitive Sciences*, *12*(2), 48–53. https://doi.org/10.1016/j.tics.2007.11.008

Sip, K. E., Smith, D. V., Porcelli, A. J., Kar, K., & Delgado, M. R. (2015). Social closeness and feedback modulate susceptibility to the framing effect. *Social Neuroscience*, *10*(1), 35–45. https://doi.org/10.1080/17470919.2014.944316

Spence, S. a., Kaylor-Hughes, C., Farrow, T. F. D., & Wilkinson, I. D. (2008). Speaking of secrets and lies: The contribution of ventrolateral prefrontal cortex to vocal deception. *NeuroImage*, *40*(3), 1411–1418. https://doi.org/10.1016/j.neuroimage.2008.01.035

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*(5811), 515–518. https://doi.org/10.1126/science.1134239

Vanberg, C. (2008). Why Do People Keep Their Promises? An Experimental Test of Two Explanations1. *Econometrica*, *76*(6), 1467–1480. https://doi.org/10.3982/ECTA7673

Volz, K. G., Vogeley, K., Tittgemeyer, M., Cramon, D. Y. von, & Sutter, M. (2015).

The neural basis of deception in strategic interactions. *Frontiers in Behavioral Neuroscience*, *9*(February), 1–12. https://doi.org/10.3389/fnbeh.2015.00027

Vrij, a., Granhag, P. a., & Porter, S. (2010). Pitfalls and Opportunities in Nonverbal and Verbal Lie Detection. *Psychological Science in the Public Interest*, *11*(3), 89–121. https://doi.org/10.1177/1529100610390861

Wiley, S. D. (1998). Deception and detection in psychiatric diagnosis. *The Psychiatric Clinics of North America*, *21*(4), 869–893.

Yang, Y., Raine, A., Lencz, T., Bihrle, S., Lacasse, L., & Colletti, P. (2005). Prefrontal white matter in pathological liars. *British Journal of Psychiatry*. https://doi.org/10.1192/bjp.187.4.320

Yang, Y., Raine, A., Narr, K. L., Lencz, T., LaCasse, L., Colletti, P., & Toga, A. W. (2007). Localisation of increased prefrontal white matter in pathological liars. *British Journal of Psychiatry*. https://doi.org/10.1192/bjp.bp.106.025056

Yates, G. P., & Feldman, M. D. (2016). Factitious disorder: A systematic review of 455 cases in the professional literature. *General Hospital Psychiatry*, *41*, 20–28. https://doi.org/10.1016/j.genhosppsych.2016.05.002

Zhou, Y., Wang, Y., Rao, L. L., Yang, L. Q., & Li, S. (2014). Money talks: Neural substrate of modulation of fairness by monetary incentives. *Frontiers in Behavioral Neuroscience*. https://doi.org/10.3389/fnbeh.2014.00150