¿PORQUÉ RAZONAN LOS HUMANOS? ARGUMENTOS PARA UNA TEORÍA ARGUMENTATIVA[1]

Hugo Mercier
Institut Jean Nicod
hugo.mercier@gmail.com

Dan Sperber
Institut Jean Nicod
sperberd@ceu.edu

Traducción:
Juan Manuel Vivas
Universidad Nacional de Rosario
jmndrx@gmail.com

Cecilia McDonnell
Universidad Nacional de Rosario
cecilia.mcd@gmail.com

1 Versión Original: Mercier, Hugo, y Dan Sperber. «Why Do Humans Reason? Arguments for an Argumentative Theory». *Behavioral and Brain Sciences* 34.02 (2011): 57-74. La presente traducción cuenta con la autorización expresa de los autores y de Cambridge University Press.

La *inferencia* (tal como el término se entiende generalmente en psicología) es la producción de nuevas representaciones mentales sobre la base de representaciones previas. Ejemplos de inferencias son la producción de creencias nuevas asentadas en anteriores, de expectativas sobre base de la percepción, o de planes fundados en preferencias y creencias. Así entendida, la inferencia no tiene por qué ser deliberada o consciente, y sucede no sólo en el pensamiento conceptual, sino también en la percepción y en el control motriz (Kersten et al., 2004, Wolpert & Kawato 1998). Es un ingrediente básico de cualquier sistema cognitivo. El *razonamiento*, tal como es entendido comúnmente, se refiere a una forma muy especial de inferencia a nivel conceptual, donde no solo se produce una nueva representación mental (o conclusión) de manera consciente, sino que también se figuran del mismo modo representaciones sostenidas previamente (o premisas) que la garantizan. Éstas son interpretadas como las responsables de proveer las razones que permiten aceptar la conclusión. La mayor parte del trabajo en la psicología del razonamiento se vincula con éste entendiéndolo de la manera expuesta, debido a que tal tipo de razonamiento es típicamente humano. No existe evidencia de que esto ocurra en animales no humanos o en niños pre-verbales[2].

¿Cómo razonan los seres humanos? ¿Por qué razonan? Estas dos preguntas son mutuamente relevantes, dado que los mecanismos del razonamiento deberían hallarse ajustados a su función. Mientras que la pregunta acerca del *cómo razonamos* ha sido investigada de forma sistemática (por ejemplo, Evans et al. 1993; Johnson-Laird 2006; Oaksford & Chater 2007; Rips 1994), la pregunta acerca de *por qué razonamos* ha sido muy poco discutida. ¿A qué se debe esto? Quizás a que la función del razonamiento se considera harto obvia como para merecer demasiada atención. De acuerdo a una larga tradición filosófica, el razonamiento es lo que permite a la mente humana ir más allá de la mera percepción, del hábito y del instinto. En la primera sección de este artículo (la parte teórica) esbozamos la pregunta acerca de "cómo" razonamos, para luego

2 Recientemente, "razonamiento" ha sido utilizado como un simple sinónimo de "inferencia" y, de este modo, es atribuido acríticamente a los infantes (Spelke & Kinzler 2007) o a animales no humanos (Blaisdell et al. 2006). En este artículo, sin embargo, usamos "razonamiento" en su uso más común y llano. El contenido de este artículo debería dejar en claro por qué vemos esto como una elección de principios terminológicos.

concentrarnos en la pregunta acerca del "por qué": trazamos un abordaje basado en la idea de que la función primaria para la cual el razonamiento evolucionó es la producción y la evaluación de argumentos en la comunicación. En las secciones 2 a 5, consideramos algunos de los temas y hallazgos principales en la literatura experimental, y mostramos cómo nuestro abordaje contribuye a comprender mejor gran parte de la evidencia experimental ganando, de este modo, apoyo empírico.

1. Razonamiento: Mecanismo y función

1.1. Inferencia y argumento intuitivos

Desde la década del'60, la psicología del razonamiento ha sugerido que los humanos razonan escasamente, por lo que no consiguen llevar a cabo tareas lógicas simples (Evans 2002), cometiendo, así, graves errores en términos probabilísticos (Kahneman & Tversky 1972; Tversky & Kahneman 1983), y quedando sujetos a diversas inclinaciones irracionales a la hora de tomar decisiones (Kahneman et al. 1982). Este hecho ha llevado a repensar los mecanismos del razonamiento, pero no –o al menos no en el mismo grado– la función asumida de mejorar el conocimiento humano y la toma de decisiones. El desarrollo más importante ha sido la emergencia de los modelos de procesos duales que distinguen intuiciones de razonamientos (o razonamiento de sistema 1 y de sistema 2) (Evans 2007; Johnson-Laird 2006; Kahneman 2003; Kahneman & Frederick 2002; 2005; Sloman 1996; Stanovich 2004). Según la aproximación que aquí proponemos, los procesos duales son aquellos en los que los argumentos utilizados en el razonamiento son el resultado de un mecanismo de inferencia intuitiva (Mercier & Sperber 2009; Sperber 1997; 2001).

Un *proceso de inferencia* es un proceso, el resultado representacional de lo que necesaria o probablemente se sigue de su origen representacional. La función de un proceso inferencial es aumentar y corregir la información que dispone el sistema cognitivo. Una aproximación evolucionista sugiere que los procesos inferenciales, antes que estar basados en un mecanismo inferencial simple o constituir un sistema integrado simple, son más bien propensos a ser desempeñados por una variedad de mecanismos de

dominios específicos, cada uno de los cuales se halla de acuerdo con las demandas específicas y las capacidades de su dominio (e.g., ver Barkow et al. 1992). El proceso inferencial llevado a cabo por estos mecanismos son inconscientes: no se trata de actos mentales que los individuos deciden realizar, sino de procesos que tienen lugar dentro de sus cerebros, en un nivel "sub-personal" (en el sentido que le otorga Dennet 1969). Las personas pueden ser conscientes de haber rechazado una conclusión determinada –es decir, el resultado de un proceso inferencial–, pero nosotros afirmamos que nunca están al tanto del proceso en sí mismo. Todas las inferencias llevadas a cabo por los mecanismos inferenciales son, en este sentido, intuitivas. Ellas generan las creencias intuitivas, esto es, opiniones sostenidas sin conocer los motivos que llevan a aseverarlas.

La afirmación según la cual son inconscientes e intuitivos todos aquellos procesos inferenciales llevados a cabo por mecanismos inferenciales especializados, puede parecer contradecir la experiencia común de la formación de una creencia, ya que se ha reflexionado en vistas a sostenerla –y no, o no únicamente, se las afirma por su fuerza intuitiva. Tales opiniones, sostenidas conscientemente, son mejor descritas como creencias reflectivas antes que como intuitivas (Sperber 1997). La razón que albergamos a nivel consciente que nos permite aceptar una creencia reflexiva puede ser confiada a su fuente (el profesor, el doctor, el cura). Nuestras razones pueden, asimismo, estar relacionadas con el contenido de la creencia: nos damos cuenta, por ejemplo, de que sería inconsistente de nuestra parte limitarnos a creencias previas y no sopesar nuevos argumentos que nos son dados. Ahora bien, lejos de negar que pudiéramos arribar a una creencia mediante la reflexión, asumimos este proceso como un razonamiento propiamente dicho (lo cual es el asunto principal de este artículo). Lo que caracteriza al razonamiento apropiado es la conciencia tanto de una conclusión determinada como de un argumento que justifica aceptar tal conclusión. Sugerimos, sin embargo, que los argumentos explotados en el razonamiento son el resultado de un mecanismo inferencial intuitivo. Como todos los demás mecanismos inferenciales, sus procesos son inconscientes (tal como ha afirmado Johnson-Laird 2006, p. 53; y Jackendoff 1996) y sus conclusiones son intuitivas. De cualquier forma, estas conclusiones intuitivas son

casi argumentos, es decir, casi representaciones de las relaciones entre premisas y conclusiones.

Las inferencias intuitivas hechas por los humanos no se refieren únicamente a objetos ordinarios y eventos en el mundo. Pueden ser, asimismo, sobre representaciones de tales objetos o eventos (o también, en un orden más elevado, representaciones de representaciones). La capacidad de realizar esto último y de esbozar inferencias sobre su base, es una capacidad meta-representacional con propiedades formales relevantes a las computaciones mentales que están involucradas (Recanati 2000; Sperber 2000b). Muchos mecanismos mentales usan esta capacidad meta-representacional. En particular, los humanos tienen un mecanismo para representar representaciones mentales y para esbozar inferencias intuitivas acerca de ellas. Este mecanismo de la Teoría de la Mente es esencial para la comprensión de los otros y de nosotros mismos (Leslie 1987; Premack & Woodruff 1978). Los humanos también cuentan con un mecanismo para representar representaciones verbales y para esbozar inferencias intuitivas acerca de ellas. Este mecanismo pragmático es esencial para la comprensión de los sentidos comunicados en contexto (Grice 1975; Sperber & Wilson 2002).

Queremos afirmar que existe incluso otro mecanismo intuitivo meta-representacional, a saber, un mecanismo que permite representar posibles razones para aceptar una conclusión –es decir, para representar argumentos– y evaluar su fuerza. Los argumentos deberían ser distinguidos de las inferencias con claridad. Una inferencia es un proceso cuyo resultado es una representación. Un argumento es una representación compleja. Tanto la inferencia como el argumento tienen lo que puede llamarse una conclusión, pero en el caso de la inferencia, la conclusión es el resultado de la inferencia; en el caso del argumento, la conclusión es una parte –típicamente la última– de la representación. El resultado de una inferencia puede ser llamado “conclusión” porque lo que caracteriza a un proceso inferencial es que su resultado está justificado en su origen; la manera, sin embargo, en la cual éste lo evidencia no se encuentra representada en el resultado de una inferencia intuitiva. Lo que hace que la conclusión de un argumento sea, de hecho, una "conclusión" (y no simplemente una proposición),

es que las razones para establecerla sobre la base de las premisas son (al menos parcialmente) explicadas. Justamente, como ha afirmado Gilbert Harman (1986), es un error común pero costoso el confundir los pasos causales y temporales de una inferencia con los pasos lógicos de un argumento. Los primeros no necesitan recapitular pasos lógicos de argumento alguno para ser una inferencia y, a su vez, los últimos no necesitan ser seguidos de alguna inferencia para ser un argumento.

El famoso argumento del *cogito* cartesiano "pienso, luego existo" ilustra la manera en que un argumento puede ser el resultado de una inferencia intuitiva. La mayoría de la gente cree intuitivamente que efectivamente existe y no está buscando razones para justificar esa creencia. Pero si uno tuviera que buscar tales razones –es decir, tomar una actitud reflexiva en torno a la proposición de que uno existe– el argumento de Descartes probablemente nos hubiera convencido: es intuitivamente evidente que el hecho de que estamos pensando es una razón lo suficientemente buena como para aceptar que existimos o, en otros términos, que sería inconsistente aceptar "yo pienso" y negar "yo soy". Lo que no es obvio en absoluto es que este argumento intuitivamente bueno es verdaderamente un buen argumento, y los filósofos han estado discutiendo acaloradamente el asunto (e.g., Katz 1986).

Sean simples como el *cogito* o más complejos, todos los argumentos deben, en último término, estar cimentados en juicios intuitivos que, dadas ciertas conclusiones, se sigan de determinadas premisas. En otras palabras, estamos sugiriendo que los argumentos no son el resultado de un mecanismo de sistema 2 por razonamiento explícito, que estaría apartado de, y en simétrico contraste con, un mecanismo de sistema 1, por inferencia intuitiva. Antes bien, los argumentos son el resultado de un mecanismo de inferencia intuitiva entre muchos, que ofrece intuiciones acerca de las relaciones entre premisa y conclusión. Las intuiciones acerca de los argumentos tienen un componente evaluativo: algunos argumentos son vistos como fuertes, otros como débiles. Además podrían existir argumentos en competencia por conclusiones opuestas y podríamos intuitivamente preferir una por sobre la otra. Esas evaluaciones y preferencias se basan, en última instancia, en la intuición.

Si aceptamos una conclusión a causa de un argumento en su favor, que es lo suficientemente fuerte a nivel intuitivo, esta aceptación es una decisión epistémica que tomamos a nivel personal. Si construimos un argumento complejo al ir uniendo pasos argumentativos, de los cuales creemos que cada uno posee una fuerza intuitiva suficiente, esta es entonces una acción mental de nivel personal. Si producimos verbalmente el argumento de manera tal que otros vean su fuerza intuitiva y acepten su conclusión, es una acción pública a la cual nos sometemos de manera consciente. La acción mental de llevar a cabo un argumento convincente, la acción pública de producir verbalmente este argumento de manera tal que los demás queden convencidos por él, y la acción mental de evaluar y aceptar la conclusión de un argumento producido por otros, corresponden a lo que es común y tradicionalmente señalado como "razonamiento" (un término que puede referir tanto a la actividad mental como a la verbal).

¿Por qué debería la explotación reflexiva de un mecanismo de inferencia intuitiva, entre tantos, destacarse como importante y convertirse en lo que diferencia a los hombres de las bestias? ¿Por qué, en las teorías del razonamiento de doble proceso, debería ser contrastado por sí mismo a partir de todos los mecanismos de inferencia intuitiva de manera conjunta? Vemos tres explicaciones complementarias que muestran la importancia del razonamiento. En primer lugar, cuando razonamos, sabemos que estamos razonando, mientras que la mera existencia de la inferencia intuitiva fue vista de manera controversial en la filosofía, previo a su descubrimiento en la ciencia cognitiva. En segundo lugar, si bien un mecanismo inferencial que proporciona intuiciones sobre argumentos es, estrictamente hablando, altamente específico del dominio, los argumentos sobre los que proporciona intuiciones pueden ser representaciones de cualquier cosa. Así, cuando razonamos sobre la base de estas intuiciones, podemos llegar a conclusiones en todos los dominios teóricos y prácticos. En otras palabras, aunque las inferencias sobre los argumentos son específicas del dominio (como esperan los psicólogos evolucionistas), tienen consecuencias generales de dominio y proporcionan una especie de generalidad de dominio virtual (sin la cual

los enfoques tradicionales y de doble proceso tendrían poco sentido). En tercer lugar, como argumentaremos en el siguiente apartado, la función misma del razonamiento se pone en evidencia en la comunicación humana.

1.2. La función del razonamiento

Utilizamos el término "función" aquí en su sentido biológico (véase Allen et al., 1998). En pocas palabras, la función de un rasgo es un efecto del mismo que explica causalmente su evolución y persistencia en una población: gracias a este efecto, el rasgo ha estado contribuyendo a la aptitud de los organismos dotados con él. En principio, varios efectos de un rasgo pueden contribuir a la aptitud, y por lo tanto un rasgo puede tener más de una sola función. Incluso entonces puede ser posible clasificar la importancia de diferentes funciones y, en particular, identificar una función para la cual el rasgo se adapta mejor como su función principal. Por ejemplo, los pies humanos tienen la función de permitirnos correr y caminar, pero su postura plantígrada se encuentra mejor adaptada para caminar que para correr, lo cual constituye una fuerte evidencia de que caminar es su función principal (Cunningham et al., 2010). En el mismo sentido, no estamos argumentando en contra de la opinión de que nuestra capacidad de razonamiento puede tener varios efectos ventajosos, cada uno de los cuales puede haber contribuido a su selección como una importante capacidad de la mente humana. Argumentamos, sin embargo, que el razonamiento se adapta mejor a su papel en el desarrollo de la argumentación, que debe ser visto como su función principal. Ha habido algunos intentos tentativos en enfoques de proceso dual que pretenden explicar la función y la evolución del razonamiento. La opinión mayoritaria parece ser que la función principal del razonamiento es mejorar la cognición individual. Esto es expresado, por ejemplo, por Kahneman (2003, p.699), Gilbert (2002), Evans y Over (1996, p.154), Stanovich (2004, p.64) y Sloman (1996, p.18)). Esta concepción clásica del razonamiento, que se remonta a Descartes y a los filósofos griegos antiguos, se enfrenta a variados problemas que se hacen evidentes cuando sus reivindicaciones funcionales se exponen con un poco más de detalle. A veces se afirma (por ejemplo, por

Kahneman 2003) que la función perfectiva del razonamiento del sistema 2 se logra mediante la corrección de los errores en las intuiciones del sistema 1. Sin embargo, el razonamiento mismo es una fuente potencial de nuevos errores. Además, existen pruebas considerables que sostienen que cuando se aplica el razonamiento a las conclusiones de la inferencia intuitiva, se tiende a racionalizarlas en lugar de corregirlas (por ejemplo, Evans & Wason 1976). De acuerdo con otra hipótesis, el razonamiento consciente "nos da la posibilidad de lidiar con la novedad y de anticipar el futuro" (Evans & Over 1996, p.154).

Pero darle a un organismo la posibilidad de lidiar con la novedad y anticiparse al futuro resulta más una caracterización de un aprendizaje (o incluso, podría decirse, de la cognición en general) que del razonamiento. Después de todo, el aprendizaje puede definirse como “el proceso por el cual podemos utilizar el pasado y los eventos actuales para predecir el futuro” (Niv & Schoenbaum 2008, p. 265). La cuestión no es si, en alguna ocasión, el razonamiento puede ayudar a corregir errores intuitivos o adaptarnos mejor a nuevas circunstancias. Sin duda, puede. Lo que aquí interesa es hasta qué punto estos beneficios ocasionales explican los costos incurridos y, por tanto, la existencia misma del razonamiento entre los seres humanos y sus rasgos característicos.

En cualquier caso, las hipótesis evolutivas son de poca ayuda a menos que sean lo suficientemente precisas como para comprobar predicciones y explicaciones. Para establecer que el razonamiento tiene una función determinada, deberíamos poder al menos identificar los efectos propios de esa función en la misma forma en que funciona el razonamiento.

Aquí queremos explorar la idea de que la aparición del razonamiento se entiende mejor en el marco de la evolución de la comunicación humana. El razonamiento permite a las personas que intercambien argumentos que, en general, hacen más fiable la comunicación y, por tanto, más ventajosa. La principal función del razonamiento, afirmamos, es argumentativa (Sperber 2000a, 2001, véase también Billig 1996; Dessalles 2007; Kuhn 1992; Perelman & Olbrechts-Tyteca 1969; para un enfoque muy similar en el caso especial de Razonamiento moral, véase Gibbard 1990 y Haidt 2001).

Para que la comunicación sea estable, tiene que beneficiar tanto a emisores como a receptores; de lo contrario dejarían de enviar o de recibir, poniendo fin a la comunicación (Dawkins & Krebs 1978, Krebs y Dawkins 1984). Pero la estabilidad es a menudo amenazada por los emisores deshonestos que pueden beneficiarse manipulando a los receptores, infligiendo un coste demasiado alto para ellos. ¿Existe alguna manera de garantizar que la comunicación sea honesta? Algunas señales son indicadores fiables de la propia honestidad. Señales costosas tales como cornamentas de ciervo o colas de pavo real, señalan que el individuo es lo suficientemente fuerte como para pagar ese costo (Zahavi & Zahavi 1997). Decir "no soy mudo" es la prueba de que quien habla no es mudo. Sin embargo, para la mayoría de los ricos y variados contenidos que los humanos comunican entre sí, no hay señales disponibles que puedan ser prueba de la propia honestidad. Para evitar ser víctimas de desinformación, los receptores deben, por lo tanto, ejercer un cierto grado de lo que puede denominarse *vigilancia epistémica* (Sperber et al. 2010). La tarea de la vigilancia epistémica consiste en evaluar al comunicador y al contenido de los mensajes con el fin de filtrar la información comunicada.

Numerosos mecanismos psicológicos pueden contribuir a la vigilancia epistémica. Los dos más importantes son la calibración de confianza y la comprobación de coherencia. La gente calibra rutinariamente la confianza que les otorga a los diferentes emisores según sus competencias y benevolencia (Petty & Wegener 1998). Los rudimentos de confianza basados en la competencia se han demostrado en niños de 3 años (para revisiones, ver Clément 2010; Harris 2007). Se ha demostrado el desarollo de la capacidad de desconfiar de informantes malintencionados en etapas tempranas, entre las edades de 3 y 6 (Mascaro & Sperber 2009). La interpretación de la información comunicada consiste en activar un contexto de creencias previamente sostenidas y tratar de integrar lo nuevo con la información antigua. Este proceso puede traer a la luz incoherencias entre la información antigua y la recién comunicada. De esta manera ocurre cierta verificación de coherencia en el proceso de comprensión. En el momento en el que se descubre alguna incoherencia, el destinatario epistemológicamente vigilante

debe elegir entre dos alternativas. La más simple es rechazar la información, evitando así cualquier riesgo de ser engañado. Sin embargo, esto puede privar al destinatario de información oportuna y de la consiguiente ocasión de corregir o actualizar creencias previas. La segunda alternativa, más elaborada, consiste en asociar la comprobación de coherencia con la calibración de confianza y permitir una mayor precisión en el proceso de revisión de creencias. En particular, si alguien en quien realmente confiamos nos dice algo que es incoherente con nuestras creencias anteriores, una mínima revisión se vuelve inevitable: debemos revisar nuestra confianza en la fuente o en nuestras creencias anteriores. Probablemente escogeremos la revisión que restablezca la coherencia al menor costo, lo cual consistirá a menudo en aceptar la información y revisar nuestras creencias.

¿Cuáles son las opciones de un comunicador que desea comunicar determinada información que el destinatario probablemente no acepte con confianza? Una opción puede consistir en proporcionar pruebas de su fiabilidad en el asunto en cuestión (por ejemplo, si la información es sobre cuestiones de salud, podría informar al destinatario que ella es doctora). Pero, ¿y si el comunicador no está en posición de imponer su propia autoridad? En tal caso, una alternativa es intentar convencer al destinatario ofreciendo premisas en las que ya cree o a las que esté dispuesto a aceptar con confianza, y demostrar que, habiendo sido aceptadas tales premisas, sería menos coherente rechazar la conclusión que aceptarla. Esta opción consiste en producir argumentos en favor de las propias afirmaciones y en alentar al destinatario a examinar, evaluar y aceptar los argumentos presentados. Producir y evaluar argumentos es, por supuesto, un uso del razonamiento.

El razonamiento contribuye a la eficacia y a la fiabilidad de la comunicación al permitir que los comunicadores realicen sus afirmaciones y que los destinatarios evalúen los argumentos pertinentes. De esta manera aumenta tanto en cantidad como en calidad epistémica la información que los humanos son capaces de compartir. Nuestra afirmación de que el rol del razonamiento en la interacción social es su función principal, se ajusta a numerosos trabajos actuales que destacan el papel de la

sociabilidad de capacidades cognitivas únicas en los seres humanos (Byrne & Whiten 1988; Dunbar 1996; Dunbar & Shultz 2003; Hrdy 2009; Humhrey 1976; Tomasello et al. 2005; Whiten & Byrne 1997). En particular, el papel evolutivo de la cooperación en pequeños grupos ha sido recientemente enfatizada (Dubreuil 2010; Sterelny, en prensa). La comunicación juega un rol obvio en la cooperación humana tanto en el establecimiento de objetivos como en la asignación de deberes y derechos. La argumentación sólo es eficaz para superar los desacuerdos que probablemente se produzcan, en particular en grupos de condiciones relativamente igualitarias. Aunque difícilmente pueda haber evidencia arqueológica para la afirmación de que la argumentación ya jugó un papel importante en los primeros grupos humanos, notamos que algunos antropólogos han observado repetidamente a personas discutiendo en sociedades tradicionales de pequeña escala (Boehm et al., 1996; Brown 1991; Mercier, en prensa a).

La función principal del razonamiento es argumentativa: el razonamiento ha evolucionado y persiste principalmente porque hace que la comunicación humana sea más efectiva y ventajosa. Como la mayoría de las hipótesis evolutivas, esta afirmación corre el riesgo de ser percibida como otra "historia". Por lo tanto, es crucial demostrar que implica predicciones falsificables. Si la función principal del razonamiento es de hecho argumentativa, entonces debe exhibir como marca registrada las fortalezas y debilidades relacionadas a la importancia relativa de esta función comparadas con otras potenciales funciones del razonamiento. Esto debe comprobarse a través del trabajo experimental llevado a cabo aquí y ahora. Nuestro objetivo es explicar los efectos propios que predecimos, evaluar estas predicciones a la luz de los datos disponibles y analizar si ayudan a dar un mejor sentido a una serie de rompecabezas bien conocidos en la psicología del razonamiento y en la toma de decisiones. Si uno falla, por otra parte, en encontrar ese efecto propio de la hipotética función argumentativa del razonamiento y, aún más, si uno descubre que las principales características del

razonamiento coinciden con otra función, entonces nuestra hipótesis debería ser considerada falseada[3].

Varias predicciones pueden derivarse de la teoría argumentativa del razonamiento. La primera y más directa es que el razonamiento debe hacer bien aquello para lo cual evolucionó; esto es, producir y evaluar argumentos (secciones 2.1 y 2.2). En general, las adaptaciones funcionan mejor cuando se las utiliza para realizar la tarea para la cual se han desarrollado. En consecuencia, el razonamiento debe producir sus mejores resultados en contextos argumentativos, sobre todo en discusiones grupales (Artículo 2.3). Cuando buscamos convencer a un interlocutor de un punto de vista diferente al suyo, debemos hallar argumentos que apoyen nuestra perspectiva y no la suya. Por lo tanto, la siguiente predicción es que el razonamiento utilizado para producir el argumento debe mostrar una fuerte inclinación hacia la confirmación (sección 3). Otra predicción relacionada es que, cuando las personas razonan por su cuenta acerca de alguna de sus opiniones, es probable que lo hagan de manera proactiva, es decir, anticipando un contexto dialógico y, sobre todo, intentando encontrar argumentos que apoyen su opinión. En la sección 4 revisaremos la prueba de la existencia de tales razonamientos motivados. Finalmente, queremos explorar la posibilidad de que, incluso en la toma de decisiones, la función principal del razonamiento es producir argumentos cuyo fin sea convencer a los demás en lugar de buscar la mejor decisión. Así, se predice que el razonamiento conducirá a las personas hacia las decisiones que puedan argumentar -decisiones que puedan justificar- incluso si estas decisiones no son óptimas (sección 5).

2. Habilidades Argumentativas

2.1. Comprensión y evaluación de los argumentos

3 Nuestra hipótesis funcional será testeada sin hacer referencia a mecanismos específicos (tal y como es usual en la biología evolucionista). Incluso si uno pudiera preguntar a expensas de qué se atribuye una función argumentativa al razonamiento, se sugiere o se favorece una posición algorítmica específica, lo cual no es el foco de este artículo. No existe, en todo caso, una oposición obvia entre nuestra posición funcional y las variadas posiciones algorítmicas que se han ofrecido, por ejemplo, por Evans (2007), Johnson-Laird (2006) o Rips (1994).

En esta sección, revisaremos aquellas evidencias que muestran que las personas son discursistas calificados que utilizan el razonamiento tanto para evaluar como para producir argumentos en contextos argumentativos. Esta visión, considerada en sí misma, es compatible con otras posturas acerca de las funciones principales del razonamiento. Sin embargo, esta evidencia se torna relevante porque la idea según la cual las personas no se muestran como oradores muy competentes es relativamente común; si fuera verdad, entonces la teoría argumentativa sería imposible. Por lo tanto, es crucial demostrar que este no es el caso y que la gente cuenta efectivamente con buenas habilidades argumentativas, empezando por la capacidad de entender y evaluar los argumentos.

La comprensión de los argumentos ha sido estudiada en dos campos principales de la psicología: la persuasión y el cambio de actitud, por un lado, y el razonamiento, por el otro. Los objetivos, métodos y resultados son diferentes en cada área. En el marco de la psicología social, el estudio del primer campo ha examinado los efectos de los argumentos sobre las actitudes. En un experimento típico, los participantes escuchan o leen un argumento (un "mensaje persuasivo") y se mide la evolución de su actitud sobre el tema en cuestión. Por ejemplo, en un estudio clásico de Petty y Cacioppo (1979), se presentó a los participantes determinados argumentos en apoyo de la introducción de un examen completo de nivel superior. Algunos participantes escucharon argumentos fuertes (como datos que muestran que "las escuelas de grado y los profesionales tienen preferencia por los estudiantes que han pasado un examen completo"), mientras que otros oyeron argumentos mucho más débiles (como el testimonio de un graduado que sostiene que "puesto que una vez graduados debemos tomar exámenes completos, los estudiantes también deberían tomarlos"). En este experimento, se demostró que los participantes afectados de manera más directa por la implementación de un examen completo fueron más influenciados por los argumentos fuertes que por los débiles. Esto ilustra la conclusión más general derivada de esta literatura que, cuando se encuentran motivados, los participantes pueden utilizar el razonamiento para evaluar los argumentos con precisión (para una revisión, véase Petty & Wegener 1998).

La demostración de que las personas están capacitadas para evaluar argumentos parece contraponerse con los hallazgos de la psicología del razonamiento. En un típico experimento de razonamiento, se presenta a los participantes premisas y se les pide producir o evaluar una conclusión que deba seguirse lógicamente. De este modo, pueden llegar a tener que determinar lo que se deriva de premisas como "si hay una vocal en la tarjeta, entonces hay un número par en la tarjeta. No hay un número par en la tarjeta". En tales tareas, Evans (2002) reconoce que "el rendimiento lógico... es generalmente bastante pobre" (p. 981). Para dar un solo ejemplo, se encontró en una revisión que un promedio del 40% de los participantes no logra trazar una simple conclusión de *modus tollens* que era utilizada como ejemplo (si p entonces q, no-q, por lo tanto no-p) (Evans et al. 1993). Sin embargo, de acuerdo con la perspectiva aquí adoptada, en la mayoría de los casos el razonamiento debe proporcionar una evaluación valiosa en contextos dialógicos dados -cuando alguien está intentando genuinamente de convencernos de algo. Ese no es el caso en estas tareas descontextualizadas que no implican interacción ni problemas abstractos. De hecho, tan pronto como se puede dar sentido a estos problemas en un contexto argumentativo, el rendimiento mejora. Por ejemplo, los participantes pueden entender fácilmente un argumento *modus tollens* cuando sirve no solamente para pasar una prueba, sino también para evaluar información comunicada (véase Thompson et al 2005b.); la producción de argumentos válidos de *modus tollens* en contextos argumentativos resulta así "sorprendentemente común" (Pennington y Hastie 1993, p. 155).

Así como algunos que se dedican al estudio del razonamiento centrándose en falacias lógicas, otros investigadores se han dedicado al estudio de las falacias en la argumentación. A diferencia de las primeras, estas últimas se presentan en grados: en función de su contenido y su contexto, pueden ser más o menos falaces. Por ejemplo, una falacia de pendiente resbaladiza (donde la afirmación se critica por ser un paso en una pendiente que termina en un flagrante error) es, de hecho, válida en la medida en que, habiendo dado el primer paso en la pendiente, es probable que uno continúe hasta el final (Corner et al., 2006).

Varios experimentos han demostrado que los participantes son generalmente capaces de detectar otras falacias argumentativas (Hahn & Oaksford 2007, el experimento 3; Neuman 2003; Neuman et al. 2006; Weinstock et al. 2004; ver también Corner & Hahn 2009). No sólo las detectan, sino que además tienden a reaccionar de forma apropiada: rechazándolas cuando son de hecho erróneas, o convenciéndose si se encuentran bien fundadas (Corner et al., 2006; Hahn & Oaksford 2007; Hahn et al. 2005; Oaksford & Hahn 2004; Rips 2002). Cuando los investigadores estudiaron otras habilidades específicas de la argumentación, el rendimiento ha demostrado ser satisfactorio. Así, los participantes resultan capaces de reconocer la macroestructura de los argumentos (Ricco 2003), de seguir los compromisos de los diferentes oradores (Rips 1998), y de atribuir la carga de la prueba apropiada (Bailenson y Rips 1996; véase también Rips 1998, experimento 3). En conjunto, los resultados revisados en esta sección demuestran que las personas son buenas para evaluar argumentos tanto en el nivel de las inferencias individuales como en el de las discusiones completas.

2.2. Producir argumentos

Los primeros estudios que investigaron sistemáticamente la producción de argumentos utilizó la siguiente metodología[4]: a los participantes se les solicitó que pensaran en un tema determinado, como por ejemplo "¿restaurar el servicio militar aumentó significativamente la capacidad de EEUU de influir en los acontecimientos del mundo?" (Perkins, 1985) o "¿Cuáles son las causas del fracaso escolar?" (Kuhn, 1991); después de reflexionar durante unos minutos, tuvieron que definir y defender su visión frente al experimentador. Las conclusiones de estos estudios fueron bastante sombrías y destacaron tres principales defectos. El primero es que la gente recurre a meras explicaciones (teorías causales "que tienen sentido") en lugar de depender de verdadera

4 En la psicología del razonamiento, algunas tareas pueden ser descritas como *tareas de producción* porque los participantes deben producir una conclusión lógicamente válida desde un grupo de premisas. Sin embargo, estas tareas resultan muy diferentes desde la producción de argumentos en el marco de un debate. En un contexto dialógico, uno comienza con la conclusión e intenta encontrar premisas que puedan convencer al interlocutor. Este es el significado de *producción* que resulta relevante aquí.

evidencia (datos) para apoyar sus puntos de vista. Sin embargo, una investigación posterior ha demostrado que esta situación es sobre todo un efecto de la falta de evidencia a disposición de los participantes: cuando ésta se encuentra disponible, los participantes muestran una preferencia hacia ella (tanto en la producción como en la evaluación) (Brem y Rips 2000; véase también Hagler y Brem 2008; Sá et al. 2005). Un segundo defecto señalado por Perkins y Kuhn es la relativa superficialidad de los argumentos utilizados por los participantes. No obstante, esto puede ser explicado por una característica de las tareas propuestas: a diferencia de lo que sucede en un verdadero debate, el experimentador no cuestionó los argumentos de los participantes, por más débiles que fueran. En un contexto argumentativo normal, un buen argumento es un argumento que no es refutado. Mientras no sean cuestionados, tiene sentido estar satisfecho con argumentos aparentemente superficiales. Por otro lado, las personas deberían ser capaces de generar mejores argumentos cuando participan en un verdadero debate. Esto es exactamente lo que Kuhn y sus colegas observaron: los participantes que tenían que debatir un determinado tema mostraron una significativa mejora en la calidad de los argumentos que utilizaron después (Kuhn et al., 1997; para obtener resultados similares con el razonamiento analógico, consulte Blanchette y Dunbar 2001).

El tercer defecto, de acuerdo con Perkins y Kuhn, es el más relevante aquí. En general los participantes habían fracasado en anticipar contra-argumentos y generar réplicas. Para estos dos autores, y con la tradición de pensamiento crítico es su conjunto, este hecho es un defecto muy grave. Visto desde un perspectiva argumentativa, sin embargo, esto puede no ser una falla sencilla sino más bien una característica de la argumentación que contribuye a su eficacia en el cumplimiento de su función. Si la meta de uno es convencer a los demás, debe buscar principalmente argumentos que apoyen su perspectiva. Buscar contra-argumentos contra las propias afirmaciones puede ser parte de una estrategia argumentativa más sofisticada orientada a anticipar la respuesta del interlocutor pero, en el entorno experimental, no se daba un ida y vuelta que fomentara ese esfuerzo extra (y los participantes sabían que no debían esperarlo). Si ésta es una correcta explicación de lo que no tiene que ser un defecto, entonces la

dificultad que las personas parecen tener en encontrar contra-argumentos debería ser fácilmente superada al cuestionar las afirmaciones de otra persona en lugar de defender las propias opiniones. De hecho, cuando a los miembros del jurado se les pidió que llegaran a un veredicto y, luego, cuando se les presentó un veredicto alternativo, casi todos ellos fueron capaces de encontrar argumentos en su contra (Kuhn et al. 1994). En otro experimento, todos los participantes fueron capaces de encontrar contra-argumentos para una afirmación (que no era de ellos) y lo hicieron muy rápidamente (Shaw 1996).

Cuando la gente tuvo la oportunidad de observar el rendimiento en el razonamiento en condiciones argumentativas afortunadas, encontraron buenos resultados. Resnick y sus colegas (1993) crearon grupos de tres participantes que se hallaban en desacuerdo sobre una cuestión determinada. Analizando los debates, los investigadores quedaron "impresionados por la coherencia del razonamiento desplegado. Los participantes ... parecían construir complejos argumentos y estructuras de ataque. Las personas parecen ser capaces de reconocer estas estructuras y de atacar eficazmente sus componentes individuales, así como el argumento en su totalidad" (pp. 362-3; véase también Blum-Kulka et al. 2002; Hagler y Brem 2008; Stein et al. 1997; Stein et al. 1996). Vale la pena señalar que un patrón sorprendentemente similar emerge de estudios desarrollistas (ver Mercier, en prensa b).

En resumen, las personas pueden ser hábiles argumentadores, produciendo y evaluando argumentos de manera correcta. Esta buena performance está en agudo contraste con los resultados abismales encontrado en otros contextos no argumentativos, lo cual queda especialmente claro por la comparación entre el desempeño individual y el desempeño del grupo.

2.3. Razonamiento de grupo

Si las personas son expertas tanto en producir como en evaluar argumentos, y si estas habilidades se muestran con mayor facilidad en contextos argumentativos, entonces los debates deberían resultar especialmente propicios para el buen desempeño

del razonamiento. Se han estudiado muchos tipos de tareas en contextos grupales, con resultados muy diversos (para revisiones recientes[5], ver Kerr y Tindale 2004; Kerr et al. 1996). Los más relevantes aquí son aquellos relativos a la lógica o, más generalmente, a las tareas intelectivas "para las que existe una demostrable respuesta correcta dentro de un sistema conceptual verbal o matemático" (Laughlin & Ellis 1986, p. 177). En los experimentos que implican este tipo de tareas, los participantes en el grupo de condición experimental suelen comenzar por la solución de los problemas de forma individual (antes de la prueba), luego resuelven los mismos problemas en grupos de cuatro o cinco miembros (en la prueba), y luego los resuelven nuevamente de forma individual (después de la prueba), para asegurarse de que cualquier mejora no se da simplemente por seguir a otros miembros del grupo. Su rendimiento se compara con los de un grupo de control de participantes que toman las mismas pruebas, pero siempre de forma individual. Las tareas intelectivas permiten una comparación directa con los resultados que provienen de la literatura de razonamiento individual, y los resultados no son ambiguos. El esquema dominante (Davis 1973) es el de *la verdad gana*, lo que significa que, tan pronto como un participante tenga entendido el problema, será capaz de convencer a todo el grupo de que su solución es la correcta (Bonner et al. 2002; Laughlin & Ellis 1986; Stasson et al. 1991)[6]. Esta situación puede dar lugar a grandes mejoras en el rendimiento. Algunos experimentos que utilizan la tarea de selección de Wason ilustran dramáticamente este fenómeno (Moshman y Geil 1998; véase también Augustinova 2008; Maciejovsky y Budescu 2007). La tarea de selección Wason es la

5 Debería subrayarse que este registro irregular podría ser parcialmente explicado por las condiciones sumamente artificiales: en la vasta mayoría de los experimentos grupales, se le pedía a los participantes que interactuaran con gente que no conocían y con la que no se volverían a encontrar, realizando tareas que no tenían significado alguno en sus vidas, afuera del laboratorio. Cuando uno de estos factores se desarrolla de manera natural, los resultados mejoran. Los debates acerca de asuntos políticos entre legos frecuentemente conducen a una mejora en términos epistémicos (Landermore, en prensa; Mercier & Landermore, en prensa). Los grupos que trabajaban juntos que han sido utilizados, resultan mucho más eficientes (Michaelsen et. al. 1989). Y el aprendizaje colaborativo es enormemente exitoso en las escuelas (Slavin 1995).

6 Otros resultados un poco más débiles han sido obtenidos en tareas de índole inductiva (Laughlin et. al. 1991; 2002; 2003; 2006). Los debates también una forma conocida de mejorar la comprensión en muchos dominios (e.g., vid. Anderson et. al. 1996; 2001; Foot et. al. 1994; Howe 1990; Johnson & Johnson 2007; 2009; Lao & Kuhn 2002; Nussbaum 2008; Nussbaum & Sinatra 2003; Slavin 1995; Smith et. al. 2009; Tolmie et. al. 1993; van Boxtel et. al. 2000; Webb & Palinscar 1996).

tarea más ampliamente utilizada en el razonamiento, y el desempeño de los participantes es generalmente muy pobre, situándose en torno al 10% de respuestas correctas (Evans 1989; Evans et al. 1993; Johnson-Laird y Wason 1970). Sin embargo, cuando los participantes tuvieron que resolver la tarea en grupos, llegaron al 80% de respuestas correctas.

Muchas y variadas críticas pueden ser formuladas contra esta interpretación de los datos. Se podría sugerir que la persona que tiene la solución correcta simplemente se la señala a los otros, quienes la aceptan de inmediato y sin mediar argumento alguno, tal vez porque la habrían reconocido como el integrante "Inteligente" del grupo (Oaksford et al. 1999). Las transcripciones de los experimentos muestran que este no es el caso: la mayoría de los participantes sólo están dispuestos a cambiar de opinión una vez que han sido totalmente convencidos de que su respuesta inicial estaba equivocada (Por ejemplo, véase Moshman y Geil 1998; Trognon 1993). Generalmente, los experimentos demuestran que los debates son esenciales para cualquier mejora del rendimiento en contextos grupales (para una revisión y algunos datos nuevos, véase Schulz-Hardt Et al. 2006; para pruebas similares en el desarrollo y literatura de educación, ver Mercier, en prensa b). Además, en estos contextos, los participantes deciden que alguien es inteligente basado en la fuerza y la pertinencia de sus argumentos y no al revés (Littlepage y Mueller, 1997). De hecho, sería muy difícil saber quién es "inteligente" en dichos grupos –aunque la inteligencia general fuera fácilmente perceptible, ésta se correlaciona solamente con un 0,33 de éxito en la tarea de selección Wason (Stanovich & West 1998). Por último, ningún participante tenía la respuesta correcta al comenzar. Varios participantes pueden estar en parte equivocados y en parte en lo correcto, pero en muchos casos el grupo era capaz de retener colectivamente sólo las partes correctas y, de esta manera, converger en la respuesta correcta. Esto nos lleva al *efecto adicional del montaje*, en el cual la actuación del grupo es mejor que la de su mejor miembro (Blinder y Morgan 2000; Laughlin et al. 2002; 2003; 2006; Lombardelli et al. 2005; Michaelsen Et al. 1989; Sniezek & Henry 1989; Stasson et al. 1991; Tindale y Ce 2002). Una vez más, hay una sorprendente convergencia aquí con la literatura del desarrollo que

muestra cómo los grupos -incluso cuando ningún miembro tenía la respuesta correcta inicialmente- pueden facilitar el aprendizaje y la comprensión de una amplia variedad de problemas (Mercier, en prensa b).

De acuerdo con otro argumento en contra, las personas están simplemente más motivadas, por lo general, cuando se encuentran en grupos (Oaksford et al. 1999). Esto no es así[7]. Por el contrario, "el hallazgo siempre presente en muchas décadas de investigación (por ejemplo, véase Hill 1982; Steiner 1972) es que los grupos, por lo general, no cumplen con las líneas de base de productividad razonablemente potenciales" (Kerr y Tindale 2004, p. 625). Por otra parte, otros tipos de motivación no tienen ese efecto beneficioso sobre el razonamiento. Incluso cuando intervienen grandes incentivos monetarios, aunque sean sustanciales, fallan en mejorar el rendimiento en tareas de razonamiento y toma de decisiones (Ariely et al, 2009;. Bonner y Sprinkle 2002; Bonner Et al. 2000; Camerer y Hogarth 1999; y, en el específico caso de la tarea de selección de Wason, véase Johnson-Laird y Byrne 2002; Jones y Sugden, 2001). Por lo tanto, no todo incentivo será suficiente: el contexto grupal tiene un poder de motivación al cual el razonamiento responde específicamente[8].

La teoría argumentativa también ayuda a predecir lo que va pasar en contextos grupales no óptimos. Si todos los miembros del grupo comparten una opinión, no debería surgir un debate espontáneamente. Sin embargo, en muchos contextos experimentales e institucionales (jurados, comisiones), las personas se ven obligadas a hablar, incluso si ya estaban de acuerdo. Cuando todos los miembros están de acuerdo en un cierto punto de vista, cada uno de ellos puede encontrar argumentos en su favor. Estos argumentos no serán examinados críticamente, mucho menos refutados,

7 Incidentalmente, otra ventaja de la teoría sugerida aquí es que hace posible predicciones testeables acerca de los contextos que podrían motivar el uso del razonamiento; principalmente, contextos en los que toma lugar una argumentación real o anticipada. Esto contrasta con las teorías estándar de doble proceso, forma comprobable de principios de cuándo se debe activar el razonamiento del sistema 2.

8 Es posible que valga la pena mencionar que lo que la motivación general no logra es un razonamiento eficiente o libre de prejuicios, más que el razonamiento *per se*. Si usted paga a personas para tener la respuesta correcta en, por ejemplo, la tarea de selección de Wason, ellos podrían razonar más, pero aún con prejuicios, y su respuesta permanecerá incorrecta.

proporcionando así, a los otros miembros del grupo, razones adicionales para sostener esa opinión. El resultado debería ser un fortalecimiento de las opiniones sostenidas por el grupo (para una revisión, véase Sunstein 2002; para una reciente ilustración, ver Hinsz et al. 2008). Contra la ley de la polarización de grupos de Sunstein, es importante tener en cuenta que este resultado es específico de contextos artificiales en los que las personas debaten a pesar de tender a estar de acuerdo en primer lugar. Cuando los miembros del grupo no están de acuerdo, las discusiones a menudo conducen a la despolarización (Kogan y Wallach 1966; Vinokur y Burnstein 1978). En ambos casos, el comportamiento del grupo se puede predecir sobre la base de la dirección y la fuerza de los argumentos accesibles a los miembros del grupo, como se ha demostrado en las investigaciones llevadas a cabo en el marco de la teoría del argumento convincente (Vinokur 1971), que concluye con la predicción del marco actual (Ebbesen y Bowers 1974; Isenberg 1986; Kaplan & Miller 1977; Madsen 1978).

La investigación revisada en esta sección muestra que las personas son expertas argumentadoras: pueden usar el razonamiento tanto para evaluar como para producir los argumentos. Este buen desempeño ofrece un fuerte contraste con los pobres resultados obtenidos en tareas de razonamiento abstracto. Por último, la mejora del rendimiento en los entornos argumentativos observados confirma que el razonamiento se encuentra allí en su mejor momento. Ahora exploraremos con mayor profundidad un fenómeno ya mencionado en esta sección: el sesgo de confirmación.

3. El sesgo de confirmación: ¿Un defecto del razonamiento o una característica propia de la producción de argumentos?

El sesgo de confirmación consiste en la "búsqueda o interpretación de las pruebas de manera parcial en torno a creencias existentes, expectativas, o a una hipótesis a mano" (Nickerson, 1998 pág. 175). Es uno de los sesgos más estudiados de la psicología (Para una revisión, ver Nickerson, 1998). Si bien existe una variación individual, parece que todo el mundo se ve afectado en cierta medida, independientemente de factores como la inteligencia general o la apertura mental

(Stanovich & West 2007; 2008a; 2008b). Para las teorías estándar de razonamiento, el sesgo de confirmación no es más que un defecto de razonamiento. Para la teoría argumentativa, sin embargo, es una consecuencia de la función del razonamiento y por lo tanto una *característica* del razonamiento cuando se utiliza para la producción de argumentos.

De hecho, sugerimos, que la etiqueta *sesgo de confirmación* ha sido aplicada a dos tipos de casos distintos que se caracterizan por fracasar al buscar evidencias contrarias o contra-argumentos para una creencia existente y que se hallan en consonancia con la aproximación argumentativa, aunque derivados de diferentes maneras. En los casos que amerita la etiqueta de sesgo de confirmación, lo que se intenta es convencer a los demás. Por lo general se buscan argumentos y evidencias que confirmen la propia afirmación, ignorando aquellos que no lo hacen y refutando los que fueron anticipados. Si bien esto puede ser visto como un sesgo desde una perspectiva epistemológica normativa, sirve claramente al objetivo de convencer a otros. En otros casos, tratamos no con razonamientos parciales sino con la ausencia de un razonamiento apropiado. Ésta es de esperar en personas que ya cuentan con alguna creencia sobre la base de su percepción, su memoria o alguna inferencia intuitiva y, por ello, no precisan argumentar en favor de ella. Si creo que mis llaves están en mis pantalones, es porque es allí donde recuerdo haberlas guardado. Después de un tiempo, podrían estar en mi chaqueta, por ejemplo. Sin embargo, a menos que tenga alguna razón positiva para pensar lo contrario, simplemente asumo que todavía están en mis pantalones, y ni siquiera hago la inferencia (que, si estoy en lo cierto, sería válida) de que no están en mi chaqueta o en cualquiera de los otros lugares en los que, en principio, podrían estar. En tales casos, la gente suele trazar inferencias más bien positivas, en lugar de negativas, a partir de sus creencias anteriores. Estas inferencias positivas generalmente son más relevantes para poner a prueba estas creencias. Por ejemplo, es más probable conseguir evidencia concluyente que apoye que tenía razón o no al buscar las llaves en los pantalones en lugar de la chaqueta (incluso si resultan estar o no en mi chaqueta, todavía podría estar equivocado por creer que están en mis pantalones). Espontáneamente derivamos

consecuencias positivas de nuestras creencias intuitivas. Esto es sólo un uso confiado de nuestras creencias, no un sesgo de confirmación (Ver Klayman y Ha 1987).

La teoría que proponemos realiza tres amplias predicciones. La primera es que el genuino sesgo de confirmación (en oposición a la mera confianza en las propias creencias intuitivas y sus consecuencias positivas) debe darse sólo en situaciones argumentativas. La segunda es que debe ocurrir únicamente en la *producción* de argumentos. Lo racional para un sesgo de confirmación en la producción de argumentos para apoyar una determinada opinión no se extiende a la *evaluación* de argumentos realizada por un público cuyo objetivo es estar bien informado. La tercera predicción es que el sesgo de confirmación en la producción de argumentos no es un sesgo a favor de la confirmación en general y en contra de la desconfirmación en general: es un sesgo a favor de la confirmación de las propias demandas, que debería ser complementado naturalmente por un sesgo a favor de la desconfirmación de las reivindicaciones opuestas y los contra-argumentos

3.1. Probando hipótesis: sin razonamiento, no hay sesgo de razonamiento

Una de las áreas en las que el sesgo de confirmación ha sido más estudiado es el de la prueba de hipótesis, usando a menudo la regla de Wason para las tareas de descubrimiento (Wason 1960). En esta tarea se les dice a los participantes que el experimentador tiene en mente una regla para generar ternas de números y que hay que descubrirla. El experimentador comienza dando a los participantes una terna que se ajusta a la regla (2, 4, 6). A continuación los participantes, pueden pensar en una hipótesis acerca de la regla y probarla proponiendo una terna de su propia elección. El experimentador dice si esa terna se ajusta o no a la regla. Los participantes pueden repetir el procedimiento hasta sentirse listos para presentar su hipótesis acerca de la regla. El experimentador les dice si su hipótesis es cierta o no. Si no es así, pueden volver a intentarlo o darse por vencidos. Los participantes mayormente proponen ternas que se ajustan a la hipótesis que tienen en mente. Por ejemplo, si un participante ha formado la hipótesis “tres números pares en orden ascendente”, podría intentar con la

serie '8, 10, 12'. Como sostienen Klayman y Ha (1987), tal respuesta corresponde a una "estrategia de prueba positiva" de un tipo que sería muy efectivo en la mayoría de los casos. Esta estrategia no se ha adoptado en una manera reflexiva, sino que más bien, sugerimos, es la forma intuitiva de explotar las hipótesis intuitivas propias. Tal situación es análoga a la comprobación de la ubicación de las llaves: sostenemos que éstas se encuentran donde creemos que las dejamos en lugar de comprobar que no se hallan en aquellos lugares en donde no creemos que deban estar. Lo que vemos aquí, entonces, es una sólida heurística, en lugar de un sesgo.

En este caso en particular, tal heurística engaña a los participantes debido solamente a algunas características muy peculiares (diseñadas expresamente) de la tarea. Lo que es realmente sorprendente es el fracaso de los intentos de conseguir que los participantes razonen con el fin de corregir sus enfoques ineficaces. Ha sido mostrado que, aun cuando se les indique que prueben falsificar las hipótesis que han generado, menos de uno de cada diez participantes es capaz de hacerlo (Poletiek 1996; Tweney et al., 1980). Dado que las hipótesis son generadas por los propios participantes, esto es lo que debemos esperar en el marco actual: la situación no es argumentativa y no activa al razonamiento. Sin embargo, si se presenta una hipótesis como procedente de otra persona, parece que más participantes tratarán de falsificarla y la abandonarán mucho más fácilmente en favor de otra hipótesis (Cowley y Byrne 2005). Lo mismo se aplica si la hipótesis es generada por un miembro de la minoría en un grupo (Butera et al. 1992). Por lo tanto, la falsificación es accesible siempre que la situación aliente a los participantes a argumentar en contra de una hipótesis que no es propia.

3.2. La tarea de selección de Wason

Una interpretación similar se puede utilizar para dar cuenta de los resultados obtenidos con la tarea de selección de Wason (Wason 1966). En esta tarea, se les da a los participantes una regla que describe cuatro Tarjetas. En la versión original, las tarjetas tienen un número en un lado y una carta en el otro, aunque sólo uno de ellos es visible -podrían ver, por ejemplo, 4, E, 7, y K. La regla podría ser, "Si hay una vocal en

un lado, entonces hay un número par en el otro lado." La tarea consiste en decir qué cartas tienen que ser dadas vuelta para determinar si la regla es verdadera. En esta tarea también es útil distinguir los efectos de los mecanismos intuitivos de los del razonamiento adecuado (como ha sido sugerido hace mucho tiempo por Wason y Evans 1975). Los mecanismos intuitivos involucrados en la comprensión de los enunciados llamará la atención del participante en relación a las cartas que se tornan más relevantes por la regla y el contexto (Girotto et al 2001;. Sperber Et al. 1995). En el caso estándar, aquéllas serán simplemente las tarjetas mencionadas en la regla (la vocal E y el número par, 4), en oposición a los que sostendrán la respuesta correcta (la E y el 7). Dado que el 4 sólo puede confirmar la regla, pero no falsearla, el comportamiento de los participantes que seleccionen esta tarjeta podría interpretarse como muestra de un sesgo de confirmación. Sin embargo, como fue primero descubierto por Evans (Evans & Lynch 1973), la simple adición de una negación en la regla ("si hay una vocal en un lado, entonces *no* hay un número par en el otro lado") deja sin modificaciones las respuestas (E y 4 siguen siendo hechos relevantes), pero en este caso las tarjetas corresponden a la respuesta correcta, que falsea la regla. Estos mecanismos intuitivos, por lo tanto, no se encuentran intrínsecamente ligados a cualquier confirmación o falsificación: simplemente señalan tarjetas que en algunos casos podrían confirmar la regla y, en otros, falsearla.

El sesgo de confirmación también se produce en la tarea de selección, pero a otro nivel. Una vez que la atención de los participantes ha sido atraída hacia algunas de las tarjetas, y han llegado a una respuesta intuitiva a la pregunta, el razonamiento no se utiliza para evaluar y corregir su intuición inicial, sino para encontrar justificaciones de la misma (Evans, 1996; Lucas & Ball 2005; Roberts & Newton 2001). Se trata de un sesgo de confirmación genuino. Al igual que con la prueba de hipótesis, esta situación no significa que los participantes simplemente sean incapaces de entender la tarea o de falsear la regla, sino que falta una motivación argumentativa adecuada. Ha quedado demostrado que los participantes pueden comprender la tarea por el buen desempeño que desarrollan en entornos grupales, tal y como lo mencionamos anteriormente. Los

participantes, por lo tanto, también deberían ser capaces de falsificar la regla cuando su primera intuición les indica que es falsa, por lo que buscan demostrarlo. Los investigadores han utilizado reglas tales como "todos los miembros del grupo A son Y" donde Y es un estereotipo negativo o positivo (Dawson Et al. 2002). Los participantes más motivados para probar la regla equivocada -los pertenecientes al grupo A cuando Y fue negativo- fueron capaces de producir más del 50% de respuestas correctas, mientras que los participantes de todas las demás condiciones (grupos distintos del 'A' y/o de estereotipo positivo) se mantuvieron debajo del 20%.

3.3. Silogismos categóricos

Los silogismos categóricos son uno de los tipos de razonamiento más estudiados. Aquí hay un ejemplo típico: "Los no-C son B; Todos los B son A; por lo tanto, algunos A no son C". A pesar de que pueden resolverse mediante programas muy simples (por ejemplo, véase Geurts 2003), los silogismos pueden resultar muy difíciles de entender - el que se acaba de ofrecer a modo de ilustración, por ejemplo, es resuelto por menos del 10% de participantes (Chater y Oaksford 1999). En términos de la teoría de modelos mentales, lo que los participantes están haciendo es construir un modelo de las premisas y derivar una posible conclusión de él (Evans et. al 1999). Esto constituye la intuición inicial de los participantes. Para resolver correctamente el problema, entonces, los participantes deberían tratar de construir contra-ejemplos a esta conclusión inicial. Pero esto significaría tratar de falsear su propia conclusión. La presente teoría predice que no es algo que harían de forma espontánea. Y de hecho, "cualquier búsqueda de modelos de contra-ejemplos es débil … los participantes basan sus conclusiones sobre el primer modelo que se les ocurre" (Evans et al 1999, p 1505; véase también Klauer et al. 2000; Newstead et al. 1999).

Una vez más, sugerimos, esto no debe interpretarse como una muestra de la falta de capacidad de los participantes, sino como falta de motivación. Cuando efectivamente quieren demostrar que una conclusión está errada, encuentran diversas maneras de falsearla.

Esto sucede con las conclusiones normales presentadas por otras personas (Sacco y Bucciarelli 2008) o cuando a los participantes se los enfrenta con las llamadas 'conclusiones increíbles', tales como “Todos los peces son truchas”. En estos casos, los participantes intentan demostrar que las premisas conducen a la conclusión lógicamente contraria (“No todos los peces son truchas”) (Klauer et al., 2000). El hecho de que la falsación conduce a mejores respuestas en estas tareas, explica por qué los participantes desempeñan mucho mejor algunas operaciones cuando la conclusión es increíble (por ejemplo, ver Evans et al. 1983). Esto no se debe a que en este caso razonen más –ya que pasan la misma cantidad de tiempo tratando de resolver problemas con conclusiones tanto creíbles como increíbles (Thompson et al. 2003). Es sólo que la dirección que toma el razonamiento está determinada principalmente por las intuiciones iniciales de los participantes. Si han llegado a la conclusión por sí mismos, o si están de acuerdo con ella, tratan de confirmarla. Si no están de acuerdo con ella, tratan de demostrar que es errónea. En todos los casos, lo que hacen es tratar de confirmar su intuición inicial.

3.4. La rehabilitación del sesgo de confirmación

En los tres casos que acabamos de examinar –la prueba de hipótesis, la tarea de selección de Wason, y el razonamiento silogístico– se puede observar un patrón similar. Los participantes tienen intuiciones que los llevan hacia ciertas respuestas. Si se utiliza el razonamiento de alguna manera en particular, se lo hace principalmente para confirmar tales intuiciones iniciales. Esto es exactamente lo que cabe esperar de una habilidad argumentativa, por lo que estos resultados refuerzan nuestra afirmación de que la función principal del razonamiento es argumentativa. Por el contrario, si las personas fueran fácilmente capaces de abstraerse de este sesgo, o si estuvieran sometidos a él sólo en contextos argumentativos, ello constituiría evidencia en contra de nuestra teoría.

De acuerdo con una explicación más estándar del sesgo de confirmación, éste sería un efecto de las limitaciones en los recursos cognitivos y, en particular, de la memoria de trabajo (por ejemplo, Johnson-Laird, 2006). No obstante, es difícil conciliar esta explicación con el hecho de que las personas resultan muy buenas en la falsación de

aquellas proposiciones con las no están de acuerdo. En esos casos, no se ven frenadas por limitaciones de recursos, incluso cuando las tareas no resultan cognitivamente más sencillas.

Sin embargo, la idea de que el sesgo de confirmación es una característica normal del razonamiento y de que juega un papel en la producción de argumentos, puede parecer sorprendente a la luz de los pobres resultados que genera, según se ha afirmado. El conservadurismo en la ciencia es un ejemplo (ver Nickerson 1998 y referencias en el mismo). Otro es el fenómeno del pensamiento de grupo, que ha sido responsabilizado de numerosos desastres, desde el fiasco de la Bahía de Cochinos (Janis 1982) a la tragedia del transbordador Challenger (Esser y Lindoerfer 1989; Moorhead et al. 1991) (para una revisión, véase Esser 1998). En tales casos, el razonamiento no tiende a ser utilizado en su contexto normal, es decir, la resolución de un desacuerdo no se desarrolla mediante una discusión. SI uno se encuentra solo o con personas que tienen puntos de vista similares, los propios argumentos no serán evaluados críticamente. Aquí es cuando es más probable que el sesgo de confirmación conduzca a resultados pobres. Sin embargo, cuando se utiliza el razonamiento en un contexto más feliz - es decir, en los argumentos presentados por personas que están en desacuerdo, pero que tienen un interés común en la verdad - el sesgo de confirmación contribuye a una forma eficiente de *división del trabajo cognitivo*.

Cuando un grupo tiene que resolver un problema, es mucho más eficiente si cada individuo busca, ante todo, argumentos para apoyar una solución dada. En ese caso, pueden presentar estos argumentos para ser puestos a prueba por otros miembros del grupo. Este método funcionará siempre y cuando la gente pueda ser influida por buenos argumentos. Los resultados analizados en la sección 2 muestran que este es generalmente el caso. Este enfoque dialógico combinado es mucho más eficiente que uno en donde cada individuo tiene que examinar por su cuenta todas las soluciones posibles cuidadosamente[9]. Las ventajas del sesgo de confirmación son aún más

9 La técnica Delphi es un método de pronóstico que puede ser vista como un intento de confirmar lo mejor posible los prejuicios mediante la crítica de diferentes expertos a las predicciones de los otros y la justificación de las propias predicciones. Su efectividad muestra que, en un contexto apropiado, la

evidentes teniendo en cuenta que a menudo en una discusión cada participante se encuentra en una mejor posición para buscar argumentos a favor de su solución preferida (situaciones de información asimétrica). Así, las discusiones de grupo proporcionan una forma mucho más eficiente de mantener regulado el sesgo de confirmación. Por el contrario, la enseñanza de habilidades en el campo del pensamiento crítico, que supuestamente está planteada para ayudarnos a superar el sesgo de manera puramente individual, no parece dar muy buenos resultados (y Richart Perkins 2005; Willingham 2008).

Para que el sesgo de confirmación pueda jugar un papel óptimo en discusiones y en el desempeño grupal, debe estar activo sólo en la producción de argumentos y no en su evaluación. Por supuesto, en el ida y vuelta de una discusión, la producción de argumentos propios y la evaluación de los del interlocutor pueden interferir entre sí, por lo que se torna difícil de evaluar adecuadamente a los dos procesos de manera independiente. Sin embargo, la evidencia revisada en la sección 2.1, acerca del entendimiento de los argumentos, sugiere fuertemente que la gente tiende a ser más objetiva en la evaluación que en la producción. Si este no fuera el caso, el éxito del razonamiento grupal revisado en la sección 2.3 sería muy difícil de explicar.

4. Razonamiento proactivo en la formación de creencias

De acuerdo con la teoría argumentativa, el razonamiento es utilizado de forma más natural en el contexto de un intercambio de argumentos durante una discusión. Pero la gente también puede ser proactiva y anticipar situaciones en las que tengan que discutir para convencer a los demás de que sus afirmaciones son ciertas o que su acciones están justificadas. Diríamos que muchos razonamientos anticipan la necesidad de discutir. En esta sección, demostraremos que los trabajos que abordan el razonamiento motivado pueden ser reinterpretados de manera provechosa bajo esta perspectiva, y, en la siguiente sección, que lo mismo se aplica para trabajar en la elección basada en el razonamiento.

confirmación de parcialidades puede conducir a un muy buen rendimiento (Green et. al. 2007; Keeney et. al. 2001; Powell 2003; Rowe & Wright 1999; Tichy 2004).

Muchas de nuestras creencias son propensas a permanecer sin ser desafiadas porque sólo son relevantes para nosotros mismos y no las compartimos, o porque son objeto de controversia únicamente con las personas con las que interactuamos, o porque tenemos la autoridad suficiente para afirmarlas. En tanto consideramos a la mayoría de nuestras creencias -en la medida en que pensamos acerca de ellos en absoluto- no como tales, sino como piezas de conocimiento, somos conscientes de que algunas de ellas son difíciles de ser compartidas universalmente o de ser aceptadas simplemente porque las expresamos nosotros. Cuando prestamos atención a la naturaleza contenciosa de estas creencias, pensamos típicamente en ellas como opiniones. Éstas son propensas a ser desafiadas, por lo que probablemente deban ser defendidas. Por este motivo, tiene sentido buscar argumentos a favor de nuestras opiniones antes incluso de tener que defenderlas. Si la búsqueda de argumentos es exitosa, estaremos listos. Si no es así, tal vez resulte una mejor opción adoptar una posición más débil que sea más fácil de defender. Tales usos de razonamiento han sido intensivamente estudiados bajo el nombre de *razonamiento motivado*[10] (Kunda 1990; véase también Kruglanski y Freund 1983; Pyszc- zynski y Greenberg, 1987; para una revisión reciente, véase Molden y Higgins 2005).

4.1. Razonamiento motivado

Una serie de experimentos realizados por Ditto y sus colegas relacionados con el razonamiento en el contexto de un resultado médico falso, ilustran la noción de razonamiento motivado (Ditto y López 1992; Ditto et al. 1998; 2003). Los participantes tuvieron que poner un poco de saliva en una tira de papel y se les dijo que, si la tira cambiaba o no de color, dependiendo de la condición, sería la indicación de una

10 Nótese que *motivado*, o *motivación*, son usadas aquí no para referirse a una motivación consciente basada en razones, como en "voy a pensar en argumentos que apoyen esta opinión propia, en caso de que alguien me pregunte más tarde". En lugar de ello, se refiere a un proceso que influencia tanto la dirección como el desencadenante del razonamiento de un modo sumamente inconsciente. Incluso a pesar de que un abogado, por ejemplo, pueda conscientemente comenzar un razonamiento e influir en la dirección que tomará, esto constituye la excepción y no la regla. Generalmente, las personas (inclusive los abogados) tienen un control limitado sobre lo que desencadenará el razonamiento o la dirección que asuma.

deficiencia de enzimas. Los participantes, habiendo sido motivados a creer que estaban sanos, trataron de reunir argumentos para sostener esta creencia. En una versión del experimento, a los participantes se les informó acerca de la tasa de falsos positivos, los cuales variaban según las condiciones. El uso que hicieron de esta información refleja el razonamiento motivado. Cuando la tasa de falsos positivos resultó alta, los participantes que fueron motivados a rechazar la conclusión, la utilizaron para socavar la validez de la prueba. Esta misma tasa alta de falsos positivos fue descartada por los participantes que habían sido motivados a aceptar la conclusión. En otra versión del experimento, se solicitó a los participantes que mencionaran datos de su historia médica que podrían haber afectado a los resultados de la prueba, lo que les dio la oportunidad de descontar estos resultados. Los participantes motivados a rechazar la conclusión, listaron más datos de este tipo y el número se correlacionó negativamente con la evaluación de la prueba. En estos experimentos, el solo hecho de que la salud del participante se estuviera poniendo a prueba, indica que no puede darse por sentada. La fiabilidad de la prueba en sí está siendo discutida. Este experimento, y muchos otros que serán revisados en este artículo, demuestran también que el razonamiento motivado no es simplemente una expresión de deseos (una forma de pensar que, si fuera común, sería en cualquier caso bastante perjudicial en su aplicación y no sería coherente con nuestra teoría). Si los deseos afectaran de hecho directamente a las creencias de esta manera, a continuación los participantes simplemente ignorarían o descartarían la prueba. En su lugar, lo que hacen es buscar evidencia y argumentos para demostrar que están sanos o al menos razones suficientes para cuestionar el valor de la prueba.

Otros estudios han demostrado el uso del razonamiento motivado para apoyar diversas creencias que otros podrían discutir. Los participantes revisan y en ocasiones alteran sus recuerdos para preservar una imagen positiva de sí mismos (Dunning Et al. 1989; Ross et al. 1981; Sanitioso et al. 1990) o modifican sus teorías causales para defender las creencias que prefieren (Kunda 1987). Algunos, incluso, cuando se les comunica el resultado de un juego en el que perdieron una apuesta, utilizan los eventos acontecidos para explicar por qué motivo deberían haber ganado (Gilovich 1983). Los

expertos políticos utilizan estrategias similares para explicar sus predicciones fallidas y reforzar sus teorías (Tetlock 1998). Los críticos, por su parte, son víctimas del razonamiento motivado y buscan defectos en un documento con el fin de justificar su rechazo cuando no están de acuerdo con las conclusiones provistas (Koehler 1993; Mahoney 1977). En contextos económicos, la gente utiliza la información de forma flexible con el fin de poder justificar las conclusiones que prefieren o de llegar a la decisión que consideran mejor (Boiney et al 1997;. Hsee 1995; 1996a; Schweitzer y Hsee 2002).

Todos estos experimentos demuestran que las personas a menudo buscan razones para justificar una opinión que están deseosos de mantener. Desde el punto de vista argumentativo, lo hacen no para convencerse de la verdad de su opinión, sino para estar listos para enfrentar los retos de otros. Si encuentran que no están preparados para cumplir con tales desafíos, pueden llegar a ser reacios a expresar una opinión que no son capaces de defender y menos favorables a la opinión en sí, pero este es un efecto individual indirecto de un esfuerzo dirigido a más personas. En un clásico marco de trabajo, donde el razonamiento se entiende como orientado a la consecución de beneficios epistémicos, el hecho de que pueda ser utilizado para justificar una opinión que ya se posee es difícil de explicar, sobre todo porque, como ahora mostraremos, el razonamiento motivado puede tener graves consecuencias epistémicas.

4.2. Consecuencias del razonamiento motivado

4.2.1. Evaluación parcial y comportamiento polarizado

En un experimento de alto renombre, Lord y sus colegas (1979) pidieron a participantes preseleccionados para defender o contravenir la pena de muerte que evaluaran los estudios relativos a su eficacia como elemento disuasorio. Los estudios otorgados a los participantes tenían diferentes conclusiones: mientras que uno parecía demostrar que la pena de muerte tenía un efecto disuasorio significativo, el otro arrojaba el resultado opuesto. A pesar de que las metodologías implementadas en ambos estudios habían sido casi idénticas, los que arrojaban una conclusión distinta a las opiniones de

los participantes fueron constantemente calificados negativamente en cuanto a su realización. En este caso, los participantes utilizaron el razonamiento no tanto para evaluar los estudios de manera objetiva, sino más bien para confirmar sus puntos de vista iniciales, intentando encontrar defectos o puntos fuertes en estudios similares en función de su conclusión. Este fenómeno se conoce como asimilación sesgada o evaluación parcial. La segunda descripción es algo engañosa. En este experimento -y en muchos de los que le han seguido- se les había pedido a los participantes que evaluaran una discusión. A pesar de ello, lo que hicieron principalmente fue producir argumentos para apoyar o refutar aquel que estaban evaluando a partir de la concordancia o no con la propia conclusión. Los participantes no intentaban formar una opinión: ya tenían una. Su objetivo es argumentativo en lugar de epistémico, y termina siendo perseguido a expensas de la solidez epistémica. Que los participantes se involucraron en esta búsqueda sesgada, incluso cuando su tarea era la de evaluar un argumento, ha quedado demostrado por los experimentos que ahora describimos.

Otros experimentos han estudiado la forma en que las personas evalúan los argumentos en función de si están de acuerdo o en desacuerdo con las conclusiones. Cuando sucede lo primero, por lo general pasan más tiempo evaluándolo (Edwards y Smith, 1996). Esta asimetría surge del hecho trivial de que rechazar lo que se nos dice suele requerir alguna justificación, mientras que no sucede lo mismo al aceptarla. Por otra parte, el tiempo dedicado a estos argumentos apunta principalmente a la búsqueda de contra-argumentos (Edwards y Smith, 1996; véase también Brock 1967; Cacioppo y Petty 1979; Eagly et al. 2000). Los participantes tienden a repasarlos para encontrar defectos y terminan, de hecho, hallando algunos, así sean problemas en el diseño de un estudio científico (Klaczynski y Gordon 1996b; Klac- zynski y Narasimham 1998; Klaczynski y Robinson 2000) o en una pieza de razonamiento estadístico (Klac- zynski y Gordon 1996a; Klaczynski y Lavallee 2005; Klac- zynski et al. 1997), o, más puntualmente, falacias argumentativas (Klaczynski 1997). En todos estos casos, el razonamiento motivado conduce a una evaluación sesgada: los argumentos con

conclusiones desfavorecidas son calificados como menos seguros y menos convincentes que los argumentos con conclusiones favorecidas.

En ocasiones, la evaluación de un argumento se encuentra sesgada hasta el punto en el que tiene un efecto opuesto al pretendido por el argumentador: al leer una discusión que deriva en una conclusión incongruente (es decir, que va en contra de las propias creencias o preferencias), los interlocutores pueden encontrar tantos defectos y contra-argumentos que su actitud inicial desfavorable, de hecho, se fortalece. Este es el fenómeno de la polarización de las actitudes, que ha sido estudiado ampliamente desde que se demostró por primera vez (Lord et al 1979; véase también Greenwald 1969; Pomerantz Et al. 1995)[11]. Taber and Lodge (2006) han demostrado que, en el campo de la política, la actitud de la polarización es más fácil de observar en los participantes que están mejor informados (véase también Braman 2009; Redlawsk 2002). Su conocimiento hace posible que estos participantes puedan encontrar más argumentos en contra, lo que lleva a evaluaciones más sesgadas.

4.2.2. Polarización, refuerzo y exceso de confianza.

La polarización de actitudes también se puede dar en circunstancias más sencillas. El sólo pensar acerca de un objeto puede ser suficiente para fortalecer las opiniones que se tengan de él (polarización). Este fenómeno se ha demostrado en numerosas ocasiones. Sadler y Tesser (1973) hicieron escuchar a los participantes de una prueba la grabación de una persona con un tono muy agradable o muy desagradable. Más tarde tuvieron que dar su opinión acerca de esta persona, después de haber pensado en ella o de haber realizado tareas de distracción. Como se esperaba, las opiniones eran más extremas (en ambas direcciones) cuando los participantes tenían que pensar en la persona. Tesser y Conlee (1975) mostraron que la polarización se incrementa con el tiempo dedicado a pensar en un elemento y Jellison y Mills (1969) demostraron que

11 La polarización de actitudes suele ocurrir en individuos que tienen una actitud muy fuerte con un alto grado de confianza. El problema es, entonces, que estos individuos tenderán a caer en un extremo de la escala de actitud antes de leer los argumentos, lo que hace casi imposible detectar cualquier movimiento hacia una actitud más extrema. Esto puede explicar, al menos en parte, las fallidas objeciones de Kuhn y Lao (1996) y Miller et. al. (1993).

aumenta con la motivación de pensar. Tal y como se mostró en el caso de la polarización que se sigue de una evaluación sesgada, aquélla se produce sólo cuando los participantes están informados (Tesser y Leona 1977; véase también Millar y Tesser 1986). El efecto puede ser mitigado al proporcionar un chequeo de datos: la simple presencia del objeto utilizado disminuirá drásticamente la polarización (Tesser 1976).

Algunos experimentos posteriores utilizaron una metodología ligeramente diferente (Chaiken y Yates 1985; Liberman y Chaiken 1991). En lugar de simplemente pensar en el objeto escogido, los participantes tenían que escribir un pequeño ensayo acerca de él. No solo se observó la polarización en este caso, sino que también se correlacionó con la dirección y el número de los argumentos ofrecidos. Estos resultados demuestran que el razonamiento contribuye a la polarización de las actitudes y sugieren fuertemente que puede ser su factor principal. Cuando se pide a la gente que piense en un objeto específico, hacia el que tienen intuitivamente una actitud positiva o negativa, lo que sucede, sugerimos, es que reflexionan menos sobre el propio objeto que en la forma de defender su actitud inicial. Muchos otros experimentos han demostrado que, una vez que las personas han formado una actitud frente a un objetivo, buscarán información que apoye tal actitud (un fenómeno conocido como exposición selectiva; ver Hart et al. 2009; Smith et al. 2008) y tratarán de usar con el mismo fin cualquier información que se les dé (Bond et al 2007;. Brownstein 2003), lo cual los lleva a elegir alternativas inferiores (Russo et al., 2006).

De acuerdo con la teoría de la argumentación, el razonamiento debería ser aún más sesgado una vez que el razonador ya ha expresado su opinión, lo que aumenta la presión para justificarla en lugar de abandonarla. Este fenómeno se llama *refuerzo* (McGuire 1964). Así, cuando los participantes se comprometen con una opinión, pensar en ella dará lugar a una polarización mucho más fuerte (Lambert et al 1996;. Millar y Tesser 1986). La *Accountability* (la necesidad de justificar las decisiones de uno) también aumentará este *refuerzo* (Tetlock et al 1989; para revisión, véase Lerner y Tetlock 1999).

Por último, el razonamiento motivado también debería afectar la confianza. Cuando los participantes piensan en una respuesta para una pregunta determinada, estarán

espontáneamente tentados a generar razones que la apoyen. Esto puede provocar, así, que se confíen demasiado en ella. Koriat y sus colegas (1980) han puesto a prueba esta hipótesis mediante el uso de preguntas de cultura general como "los sabinos eran parte de (a) la antigua India o (b) la antigua Roma". Luego de responder la pregunta, los participantes tenían que producir razonamientos que se adecuaran a sus respuestas. Se pidió a algunos que generen razones para apoyar la propia, mientras que a otros se les solicitó razones en contra de ella. Los resultados de las personas a las que se les había pedido explícitamente que generaran razones apoyando su respuesta no tuvieron diferencias con aquellos que estaban en el grupo de control, los cuales no debían aportar razones para apoyarla. Esto sugiere que lo que la gente hace de forma espontánea es pensar en razones para apoyar su propia respuesta cuando la consideran no como una pieza obvia de conocimiento, pero sí como una opinión que podría ser desafiada. Por el contrario, los participantes del otro grupo tenían mucha menos confianza. Tener que pensar en argumentos en contra de su respuesta les permitió ver sus propias limitaciones -algo que no harían por su propia cuenta (por réplicas y extensiones al fenómeno del sesgo retrospectivo y el error de atribución fundamental, ver Arkes et al. 1988; Davies 1992; Griffin y Dunning 1990; Hirt y Markman 1995; Hoch 1985; Yates Et al. 1992). Es, entonces, fácil ver que el exceso de confianza también se reduciría al hacer que los participantes discutieran sus respuestas con personas que están a favor de conclusiones diferentes.

4.2.3. Perseverancia en las creencias

El razonamiento motivado también puede ser utilizado para aferrarse a las creencias, incluso cuando han resultado ser infundadas. Este fenómeno, conocido como perseverancia en las creencias, es "uno de los fenómenos más fiables de la psicología social" (Guenther y Alicke 2008, p 706.; para una demostración temprana, véase Ross et al. 1975). La participación del razonamiento motivado en este efecto puede ser demostrada al proporcionar evidencia a los participantes tanto a favor como en contra de una creencia favorecida. Si la perseverancia en las creencias fuera el simple resultado

de un cierto grado de inercia psicológica, entonces la primera evidencia que se presentase debería ser la más influyente, ya sea que admita o no la creencia favorecida. Por otro lado, si la evidencia puede ser utilizada selectivamente, entonces debería mantenerse sólo la evidencia que apoya a la creencia favorecida, independientemente del orden de presentación. Guenther y Alicke probaron en 2008 esta hipótesis de la manera que detallamos a continuación: primero, los participantes tuvieron que realizar una tarea perceptiva simple. Esta tarea, sin embargo, fue descrita como prueba de "agudeza mental", un constructo inventado que se suponía relacionado con la inteligencia general, por lo que los resultados de la prueba asumían una gran importancia para el autoestima del participante. A continuación, a los participantes se les hizo una devolución negativa o positiva. Unos minutos más tarde, no obstante, se les dijo que ésta era en realidad falsa y se les explicó el verdadero objetivo del experimento. En tres instancias diferentes, los participantes tenían que evaluar su rendimiento: habiendo terminado la tarea, después de las votaciones, y una vez revelado el verdadero objetivo del experimento. En línea con los resultados anteriores, los participantes que habían recibido comentarios positivos mostraron un clásico efecto de perseverancia en su creencia y desestimaron la revelación, lo que les permitió preservar una imagen positiva de su eficacia. Por el contrario, aquellos que tuvieron una devolución negativa hicieron lo contrario: tomaron en cuenta la revelación, lo que les permitió rechazar la devolución negativa y restaurar una visión positiva de ellos mismos. Esto sugiere fuertemente que el tipo de perseverancia en la creencia, que acabamos de describir, es una instancia de razonamiento motivado (para aplicaciones en el campo de las creencias políticas, ver Prasad et al. 2009)[12].

4.2.4. Violación de las normas morales

Los datos revisados hasta el momento han demostrado que el razonamiento motivado puede conducir a resultados epistémicos pobres. Ahora vamos a ver que

12 Incidentalmente, esto no explica todas las formas de perseverancia en las creencias: otros mecanismos pueden ser incluidos en algunas instancias (e.g., vid. Anderson et. al. 1980), pero la disponibilidad de argumentos que apoyen la creencia desacreditada puede seguir siendo crucial (vid. Anderson et. al. 1985)

nuestra capacidad de "encontrar o inventar una razón para todo lo que uno quiera hacer" (Franklin 1799) también puede permitirnos violar incluso las propias intuiciones morales y comportarnos de manera injusta. En un reciente experimento, Valdesolo y DeSteno (2008) han mostrado el rol que el razonamiento puede jugar en el sostenimiento de la hipocresía (cuando, por ejemplo, juzgamos la acción de otro a partir de criterios morales mucho más fuertes que los que usamos para juzgar la propia). A continuación, describiremos la configuración básica utilizada. Al llegar al laboratorio, a los participantes se les asignó una de dos tareas: una corta y divertida o una larga y difícil. Por otra parte, se les dio la posibilidad de elegir qué tarea preferían realizar, sabiendo que la restante sería asignada a otro participante. También contaban con la opción de dejar que un ordenador eligiera al azar la distribución de las tareas. Una vez asignadas, los participantes tenían que evaluar qué tan justos habían sido en su elección. Otros participantes, en lugar de tener que elegir, recibían su tarea y no tenían opción: debían evaluar la justicia (fairness) con la que el participante que había hecho la tarea, conociendo exactamente las condiciones bajo las cuales la había desarrollado. De este modo, se hace posible comparar las calificaciones de los participantes que se han asignado a sí mismos la tarea fácil con las clasificaciones de aquellos a quienes se les ha asignado la tarea dura. Las diferencias entre estas dos clasificaciones es una marca de hipocresía moral. A partir de ello, los autores plantearon la hipótesis de que el razonamiento fue el responsable de esta hipocresía, ya que permite a los participantes encontrar excusas para justificar su comportamiento. Ellos probaron esta hipótesis mediante un giro de las condiciones aquí presentadas: los juicios de equidad se hicieron bajo carga cognitiva, lo que provocó que el razonamiento se tornara casi imposible. Esto tuvo el resultado pronosticado: sin la oportunidad de razonar, la calificaciones eran idénticas y no mostraron ningún indicio de hipocresía.

Este experimento es sólo un ejemplo de un fenómeno más general. El razonamiento se utiliza a menudo para encontrar justificaciones de acciones que de otra manera se percibirían como injustas o inmorales (Bandura 1990; Bandura et al 1996;. Bersoff 1999; Crandall y Eshleman 2003; Dana et al. 2007; Diekmann et al. 1997; Haidt 2001;

Mazar et al. 2008; Moore Et al. 2008; Snyder et al. 1979; para los niños, ver Gummerum et al. 2008). Tales usos del razonamiento pueden tener consecuencias terribles. Los autores de crímenes estarán tentados de "culpar a la víctima" o encontrar otras excusas para mitigar los efectos de la violación de sus intuiciones morales (Ryan 1971; para una revisión, ver Hafer y Begue 2005), lo cual puede, a su vez, facilitar la posibilidad de cometer nuevos delitos (Baumeister 1997). Este punto de vista del razonamiento encaja con recientes teorías del razonamiento moral que lo consideran como una herramienta para la comunicación y la persuasión (Gibbard 1990; Haidt 2001; Haidt y Bjorklund 2007).

Estos resultados plantean un problema a la concepción clásica del razonamiento. En todos estos casos, el razonamiento no conduce a creencias más precisas acerca de un objeto, ni a la búsqueda de respuestas correctas o a juicios morales superiores. En su lugar, al sólo buscar argumentos que apoyen la propia opinión, el razonamiento la fortalece, distorsiona las estimaciones y permite realizar violaciones a las propias intuiciones morales. En estos casos, las metas epistémicas o morales no son bien atendidas por el razonamiento. Por el contrario, el objetivo argumentativo es un aumento en la capacidad de apoyar posiciones asumidas o de justificar juicios morales.

5. Razonamiento proactivo en la toma de decisiones

En la sección anterior, hemos argumentado que buena parte del razonamiento se realiza anticipando situaciones en las que puede que tengamos que defender una opinión, y hemos sugerido que el trabajo sobre el razonamiento motivado puede ser reinterpretado fructíferamente bajo esta óptica. No sólo son las opiniones las que necesitan ser defendidas: las decisiones y las acciones que llevamos a cabo pueden requerir asimismo argumentos y, con tal fin, pueden utilizarse proactivamente razonamientos. Buscamos establecer que esta es la función principal del razonamiento en la toma de decisiones. Esta afirmación se encuentra en agudo contraste con la visión clásica que sostiene que razonar acerca de las posibles opciones y evaluar sus pros y contras, es la manera más confiable -si no la única- para llegar a decisiones sólidas

(Janis y Mann, 1977; Kahneman 2003; Simon 1955). Esta perspectiva ha sido desafiada vigorosamente en numerosas investigaciones recientes. Algunos sostienen que las mejores decisiones se basan en la intuición y se toman en fracciones de segundo (por ejemplo, véase Klein 1998), tesis que fue popularizada gracias a Gladwell (2005). Otros sostienen que la solución se halla en el inconsciente y nos aconsejan "consultarlo con la almohada" (Claxton, 1997; Dijksterhuis 2004; Dijksterhuis & van Olden 2006; Dijksterhuis et al. 2006b). Revisaremos brevemente estas críticas a la perspectiva clásica antes de pasar a considerar la literatura sobre las elecciones basadas en la razón e interpretarla a la luz de la teoría argumentativa del razonamiento.

5.1. ¿En qué medida el razonamiento ayuda a decidir?

En una primera serie de estudios, Wilson y sus colegas examinaron el efecto del razonamiento sobre la consistencia entre actitudes y comportamiento (para una revisión, véase Wilson Et al. 1989a; véase también Koole et al. 2001; Millar y Tesser 1989; Sengupta y Fitzsimons 2000; 2004; Wilson & LaFleur 1995; Wilson et al. 1984; 1989b). El paradigma básico es el siguiente: los participantes deberán expresen su actitud en relación a un objeto dado y, en una etapa, deben dar sus razones. Se ha observado consistentemente que las actitudes que se basan en razones predicen en menor grado los comportamientos futuros (y con frecuencia resultan completamente no predictivos) que las actitudes que no contaban con razones. Esta falta de correlación entre actitud y comportamiento basada en el exceso de razonamiento, puede incluso conducir a los participantes a formar preferencias intransitivas (Lee et al., 2008).

Utilizando paradigmas similares, en los que se les pide motivos a los participantes, se descubrió que el hecho de tener que proveer razones los llevó a elegir explicaciones con las que posteriormente estuvieron menos satisfechos (Wilson et al 1993), o que estaban menos en línea con las evaluaciones de los expertos (McMackin y Slovic 2000; Wilson & Schooler 1991). Los participantes, por ejemplo, empeoraron en la predicción de resultados de partidos de basquet (Halberstadt y Levine 1999). Es en línea con estos estudios que se sostiene que las personas que piensan en exceso también son propensas

a no poder entender el comportamiento de otras personas (Albrechtsen et al 2009;. Ambady y Gris 2002; Ambady et al. 2000). Esta corriente de experimentos más tarde fue seguida por Dijksterhuis y sus colegas, quienes introdujeron un paradigma modificado. En este caso, a los participantes se les proporciona una lista de características que describen diferentes artículos (tales como pisos y coches) aventajando a algunos sobre otros. En el grupo base, los participantes tenían que decir qué elemento preferían inmediatamente después de haber leído la lista. En el grupo de pensamiento consciente, se les dio unos pocos minutos para pensar en los artículos. Por último, en el grupo de pensamiento inconsciente, los participantes pasaron la misma cantidad de tiempo haciendo una tarea de distracción. A través de varios experimentos, se encontró que el mejor rendimiento se obtuvo en esta última condición: el pensamiento inconsciente fue superior al pensamiento consciente (y al de quienes tenían que decidir inmediatamente) (Dijksterhuis 2004; Dijksterhuis & van Olden 2006; Dijksterhuis et al. 2006B; 2009).

Sin embargo, algunos de los resultados de Dijksterhuis son difíciles de replicar (Acker 2008; Newell et al 2009;. Thorstein- hijo y Withrow 2009), y en algunos casos se han propuesto interpretaciones alternativas (Lassiter et al. 2009). En un meta-análisis de esta literatura, Acker (2008) ha observado que sólo en pocos experimentos fue significativamente superior el pensamiento inconsciente en relación al consciente, llegando a un valor nulo en la evaluación total de todos los experimentos. Aun así, no existía ninguna ventaja significativa del grupo de pensamiento consciente por sobre el de elección inmediata. Esta es la clase de situaciones en la que, de acuerdo con las teorías clásicas, el razonamiento debería ayudar: se tiene que tomar una nueva decisión con las opciones bien delimitadas y los pros y contras expuestos. Teniendo estas cuestiones en consideración, resulta bastante sorprendente que el razonamiento (por lo menos durante unos minutos) no nos traiga ninguna ventaja, por lo que se torna inferior a los procesos intuitivos e inconscientes. Finalmente, los estudios de toma de decisiones en ambientes naturales convergen en conclusiones similares: no sólo son la mayoría de las decisiones las que se toman de manera intuitiva, sino que cuando se utilizan

estrategias de toma de decisiones conscientes, a menudo dan lugar a resultados pobres (Klein, 1998). En la siguiente sub sección, exploramos un marco diseñado para explicar tales hallazgos al demostrar que el razonamiento no empuja a las mejores decisiones sino a las que son más fáciles de justificar.

5.2. Elección basada en la razón

A fines de 1980, un grupo de investigadores de renombre en el campo de la toma de decisiones desarrolló la estructura de la *elección basada en la razón* (para una revisión temprana, ver Shafir et al. 1993). Conforme a esta teoría, las personas a menudo toman decisiones porque pueden encontrar razones para apoyarlas. Éstas no favorecen a las mejores decisiones o a las que satisfacen criterios de racionalidad, sino a las que se puedan justificar fácilmente y tienen un menor riesgo de ser criticadas. De acuerdo con la teoría argumentativa, esto es lo que debe suceder cuando las personas se enfrentan a decisiones sobre las que sólo tienen intuiciones débiles. En este caso, el razonamiento puede ser utilizado para inclinar la balanza a favor de la opción para la cual las razones son más fáciles de alcanzar. Al menos, será más sencillo defender la decisión si el resultado muestra ser insatisfactorio.

La elección basada en la razón queda bien ilustrada en un célebre artículo de Simonson (1989) en el que estudió, en particular, el efecto de atracción (Huber et al 1982;. para una variación intercultural, ver Briley et al. 2000). El efecto de la atracción se produce cuando, dado un conjunto de dos alternativas igualmente valiosas, se añade una tercera tan buena como una de las opciones en un aspecto, pero inferior en otro. Esta adición tiende a aumentar la tasa de selección de la opción dominante en una forma no garantizada por los modelos racionales. A continuación, mostramos un ejemplo utilizado en los experimentos de Simonson. Allí, los participantes tuvieron que elegir paquetes de cerveza que variaban en precio y calidad. La cerveza A era de menor calidad que la cerveza B, pero era también más barata. Los dos atributos se equilibraban de tal manera que ambas cervezas eran elegidas regularmente en una comparación directa. Sin embargo, algunos participantes contaban, además, con la posibilidad de

elegir la cerveza C, que era más cara que la cerveza B pero no era mejor. Cuando se introdujo esta opción, los participantes tendieron a escoger la cerveza B con mayor frecuencia. Es fácil de explicar de este hallazgo en el marco de las elecciones basadas en la razón: la alternativa más pobre hace que la elección dominante sea más fácil de justificar ("La cerveza B es de la misma calidad, pero más barata que esta otra cerveza"). Para confirmar esta intuición, Simonson realizó y probó tres predicciones: (1) una elección basada en razones debe ser reforzada cuando los participantes tienen que justificarse a sí mismos, además, (2) será percibida como más fácil de justificar y menos propensa a ser criticada, y (3) debe dar lugar a explicaciones más elaboradas. Los resultados de los tres experimentos apoyaron tales predicciones. Por otra parte, estos resultados también mostraron que los participantes tendían a tomar decisiones que no se ajustaban tanto con las preferencias que habían enunciado antes de decidir. Finalmente, otra serie de experimentos demostró que, cuando los participantes pudieron utilizar más sus intuiciones, ya sea porque estaban familiarizados con las alternativas o porque las descripciones de estas alternativas estaban más detalladas, eran menos propensos al efecto de atracción (Ratneshwar Et al. 1987). Varias de las críticas más conocidas a la tesis de que los seres humanos toman decisiones racionales gracias a su capacidad de razonamiento han sido o pueden ser reinterpretadas como casos de elección basada en la razón.

6. ¿Qué es lo que la elección basada en la razón puede explicar?

6.1.1. Efecto de disyunción

El principio de seguridad (Savage 1954) afirma que cuando alguien favorece A por sobre B y ocurre el evento E pero se mantiene el mismo orden de preferencia aunque E no suceda, entonces las decisiones no resultan condicionadas por la incertidumbre acerca de la ocurrencia de E. Shafir y Tversky (1992; Tversky y Shafir 1992) grabaron varias violaciones a este principio. Por ejemplo, podemos comparar la reacción de los participantes a los siguientes problemas (Tversky y Shafir 1992):

Versiones de ganancia / pérdida

> Imagínese que usted acaba de jugar un juego de azar que le da un 50% de probabilidades de ganar $200 y un 50% de posibilidades de perder $100. La moneda se lanzó y ganó $200 o perdió $100. Ahora se le ofrece una segunda apuesta idéntica: un 50% de probabilidades de ganar $200 y un 50% de perder $100. Usted...: (a) acepta el segundo juego de azar. (B) rechaza la segunda apuesta. (Tversky y Shafir 1992, p. 306)

Ya sea que hubieran ganado o perdido en el primer juego de azar, la mayoría de los participantes aceptó la segunda apuesta. Es probable que lo hicieran por diferentes razones: en el escenario de los ganadores, razonaron que podían arriesgar sin problemas perder la mitad de los $200 que acababan de ganar. En el escenario de los que perdieron, podrían haberse tomado la apuesta como una segunda oportunidad, para compensar la pérdida. En estos dos casos, aunque la elección había sido la misma, las razones para tomarla son incompatibles. Ahora bien, cuando los participantes no saben cuál será el resultado de la primera apuesta, tienen más problemas para justificar la decisión de aceptar la segunda: las razones parecen contradecirse unas con otras. Como consecuencia, la mayoría rechaza la segunda apuesta, a pesar de que la hubieran aceptado independientemente del resultado de la primera. Los autores probaron esta explicación aún más al idear una comparación que tenía las mismas propiedades que las que acabamos de describir, excepto que las razones para "aceptar" fueron las mismas con independencia del resultado de la primera apuesta. En este caso, los participantes hicieron exactamente las mismas decisiones sabiendo o no el resultado de la primera apuesta (para un experimento similar con una variante del dilema del prisionero, ver Croson 1999).

6.1.2. Falacia del costo hundido

La falacia del costo hundido es "la tendencia a continuar una tarea una vez que un se ha hecho una inversión de dinero, esfuerzo, o tiempo" (Arkes y Blumer 1985, p. 124). Un ejemplo muy conocido de la vida real es el de Concorde: el gobierno francés y el inglés decidieron seguir pagando por un avión del que sabían que nunca obtendrían

ningún beneficio. Arkes y Ayton (1999) han argumentado que tales errores son el resultado de un uso insatisfactorio de razones explícitas como "no echar a la basura lo invertido". Realizaremos una breve revisión de la evidencia que presentan, y añadiremos más.

En primer lugar, Arkes y Ayton (1999) comparan los fuertes efectos del costo hundido observados en los seres humanos (Arkes y Blumer 1985; Garland 1990; Staw 1981) con la ausencia de este tipo de errores entre los animales[13]. También señalan que los niños no parecen propensos a este error (para evidencia más reciente y convergente, ver Klaczynski y Cottrell 2004; Morsanyi y Handley 2008). Si el razonamiento no fuera la causa sino la cura de este fenómeno, se esperaría lo contrario. Por último, algunos experimentos han variado la disponibilidad de justificaciones -un factor que no debiera ser relevante para los modelos estándar de la toma de decisiones. En este sentido, cuando los participantes pueden justificar lo perdido, son menos propensos a ser atrapados por el "costo hundido" (Soman Y Cheema 2001). Por el contrario, cuando a los participantes les resulta más difícil justificar el cambio de su curso de acción, son más propensos a caer bajo esta falacia (Bragger et al. 1998; 2003).

6.1.3. Encuadre

Los efectos de encuadre ocurren cuando las personas proporcionan diferentes respuestas a problemas estructuralmente similares que varían en su enunciación –su "marco" (Tversky y Kahneman 1981)-. Generalmente culpamos a nuestras intuiciones por estos efectos (Kahneman 2003). Otra explicación que puede ser vista como alternativa o complementaria a esta es que los diferentes marcos habilitan, en mayor o menor medida, algunas de las razones, modificando así la manera en que el razonamiento afecta nuestras decisiones. Varios resultados apoyan esta interpretación (ver McKenzie 2004; McKenzie y Nelson 2003). En primer lugar, como se ha

13 Ha sido mostrado que las palomas caen presas de la falacia, pero *solo* cuando no se dio indicación de que estaban en esa situación (Navarro & Fantino 2005). Las instrucciones recibidas por participantes humanos siempre aclara este punto de modo que los experimentos confirman el señalamiento hecho por Arkes y Ayton (1999).

mencionado anteriormente, los participantes que más razonan en torno a las tareas asignadas, son los más influenciados por los efectos de encuadre (Igou y Bless 2007). En segundo lugar, cuando los grupos toman decisiones sobre los problemas enmarcados, los grupos tienden a converger en la respuesta que se apoya en las razones más fuertes (McGuire et al 1987;. Milch et al. 2009; Paese et al. 1993). Si las respuestas de los participantes estuvieran verdaderamente basadas en sus intuiciones, las proporcionadas por el grupo tenderían a ser la media de aquéllas (Allport 1924; Farnsworth y Behner 1931). En contraposición a ello, estos resultados deben ser explicados dentro del marco de la Teoría del argumento convincente (Vinokur 1971; Vinokur y Burnstein 1978), lo que demuestra que las decisiones se basan en razones.

6.1.4. Inversión de Preferencia

La capacidad de evaluar correctamente las preferencias de cada individuo es necesaria para los modelos económicos de toma de decisiones, pero aquéllas pueden variar dramáticamente dependiendo de la forma en que se miden. Alguien puede calificar con un puntaje más elevado a A en relación a B y elegir, de todas formas, a B por encima de A (Bazerman Et al. 1992; Irwin et al. 1993; Kahneman y Ritov 1994; Slovic 1975; Tversky et al. 1988). Por ejemplo, la evaluación relativa de dos objetos puede variar o incluso ser revertida, dependiendo de si son considerados juntos o por separado (Hsee 1996b; 1998; Hsee et al., 1999). Así, cuando los siguientes objetos se presentan en forma aislada –un diccionario de música con 10.000 entradas que está "como nuevo", y uno con 20.000 entradas y una cubierta rota- la gente evalúa con mayor puntaje el de las 10.000 entradas. Sin embargo, cuando se debe elegir entre los dos, favorecen al que tiene más entradas, a pesar del estado de la cubierta (Hsee 1996b). Tales efectos se ajustan perfectamente al marco actual: la gente elige una alternativa, ya que puede ofrecer "un argumento convincente para la opción, que se puede utilizar para justificar la decisión tanto para uno mismo, como para los demás" (Tversky et al. 1988, p. 372). En el ejemplo anterior, las personas carecen de intuiciones confiables -no pueden decir cuántas entradas debería tener un buen diccionario de música-. A falta de

tales intuiciones, recaen en el razonamiento y permiten que sus juicios se guíen por la facilidad de la justificación -en este caso, la condición del diccionario que fácilmente justifica un precio alto o bajo. Por otro lado, las dimensiones con valores numéricos a menudo proporcionan justificaciones convincentes cuando las opciones se presentan de manera conjunta. Este sesgo puede conducir a decisiones menos óptimas (Hsee y Zhang 2004).

De manera más general, "quienes toman decisiones tienen una tendencia a resistir la influencia afectiva, y a depender de atributos racionales para tomar sus decisiones" (Hsee et al 2003, p 16..; véase también Okada 2005). De hecho, los atributos racionales permiten que las justificaciones sean más fáciles de enunciar. Por ejemplo, en un experimento los participantes debían elegir entre las siguientes opciones o calificarlas: un chocolate en forma de cucaracha con un peso de 56 gramos y un valor de 2 dólares, y un chocolate en forma de corazón que pesa 14 gramos y un valor de 50 centavos (Hsee 1999). Una mayoría (68%) de participantes eligieron el chocolate en forma de cucaracha, aunque más de la mitad (54%) creían que iban a disfrutar más del otro. Los participantes que eligieron el chocolate más grande, lo hizo porque el sentimiento de disgusto les parecía "irracional", difícil de justificar, en especial en comparación con la diferencia de precio y tamaño. Sin embargo, a la luz de los resultados de la psicología de disgusto (por ejemplo, Rozin et al. 1986), podemos decir que su elección fue sin duda la equivocada.

6.1.5. Otros usos inapropiados de razones

Muchos otros usos inapropiados de razones han sido empíricamente demostrados. Las decisiones de los inversores se guían por razones que parecen ser buenas, pero no están relacionadas con el rendimiento real (Barber et al. 2003). La gente utilizará una regla tal como "es mejor si hay más variedad" o "no elegir las mismas cosas que los demás" para guiar sus decisiones, incluso cuando menos variedad o más conformidad fueran de hecho más afines a sus preferencias (Ariely y Levav 2000; Berger & Heath 2007; Simonson 1990). El uso de una norma como "no pagar por los retrasos"

conducirá a comportamientos que van en contra de los propios intereses (Amir & Ariely 2003). Al pronosticar sus estados afectivos, las personas confían en teorías laicas (Igou 2004) que a menudo los llevan por mal camino (Hsee y Hastie, 2006). Porque "es mejor mantener las opciones abiertas", la gente será reacia a tomar una decisión inalterable incluso cuando sería mejor que la tomen (Gilbert & Ebert 2002). Cuando se consiente ante un acto placentero, las personas sienten que necesitan una razón para tal indulgencia, aunque ello no cambie realmente la calidad de la experiencia (Xu & Schwarz 2009). La elección basada en razones también se ha utilizado para explicar los efectos relacionados a la aversión a la pérdida (Simonson y Nowlis 2000), el efecto de equilibrar atributos (Chernev 2005), la tendencia a estar abrumados por el exceso de opciones (Scheibehenne et al., 2009; Sela et al. 2009), la característica del efecto 'creep' (Thompson et al. 2005a), el efecto de dotación (Johnson et al. 2007), aspectos de descuento temporal (Weber et al. 2007), y varias otras desviaciones de las normas de racionalidad (Shafir et al. 1993).

Otra señal de que la elección basada en la razón puede conducir a resultados no-normativos es que a veces las razones que no son relevantes para la decisión juegan, no obstante, un papel significativo. Por ejemplo, un mismo atributo irrelevante será a veces utilizado como razón para elegir un objeto (Carpenter et al. 1994) como también para no elegirlo (Simonson Et al. 1993; 1994), dependiendo de la decisión que resulte más fácil de justificar (Brown y Carpenter, 2000). Las personas también podrán verse influenciadas por informaciones irrelevantes que les resulta difícil justificar ignorarlas (Tetlock y Boettger 1989;. Tetlock et al 1996).

Todos estos experimentos demuestran usos del razonamiento poco sólidos a nivel cognitivo. Hay dos formas de explicar estos hallazgos. Se podría argumentar que se trata de casos de un mecanismo diseñado para la cognición individual y, en particular, para la toma de decisiones que a veces se utiliza del modo incorrecto. De acuerdo con la teoría argumentativa, sin embargo, la función de razonamiento es principalmente social: especialmente permite a las personas anticipar la necesidad de justificar sus decisiones frente a los demás. Este hecho señala que el uso de la razón en la toma de decisiones

debería aumentar junto con la necesidad justificarse a uno mismo. Esta predicción se ha confirmado por experimentos que muestran que la gente depende más de las razones cuando saben que sus decisiones se harán públicas más tarde (Thompson y Norton 2008) o cuando están dando consejos (en cuyo caso uno tiene que ser capaz de justificarse [ver Kray y González1999]). Por el contrario, cuando se elige por otro y no por uno mismo, se reducen estos efectos al ser menor la necesidad de una decisión utilitaria y justificable (Hamilton y Thompson 2007). Por último, cabe destacar que la imagen del razonamiento ilustrada por estos estudios puede ser excesivamente sombría: las demostraciones de que el razonamiento conduce a errores son mucho más publicables que los informes de sus éxitos (Christensen-Szalanski & Beach 1984). De hecho, en la mayoría de los casos, es probable que el razonamiento conduzca hacia buenas decisiones. Esto, sugerimos, es sobre todo porque las mejores tienden a ser más fáciles de justificar. Las razones que utilizamos para hacerlo a menudo han sido transmitidas culturalmente y es probable que apunten hacia la dirección correcta –como cuando se justifica evitar los errores del costo hundido mediante el uso de la regla que han aprendido en clase (Simonson y Nye 1992). En tales casos, las predicciones de la teoría argumentativa coinciden con las de las más clásicas. Sin embargo, lo que los resultados analizados permiten ver es que, cuando una decisión más fácilmente justificable no es buena, el razonamiento todavía nos impulsa a ella. Incluso si son casos raros, éstos resultan cruciales para comparar la teoría actual (el razonamiento nos lleva a decisiones justificables) con otras más clásicas (el razonamiento nos lleva a buenas decisiones).

7. Conclusión: el razonamiento y la racionalidad

El razonamiento contribuye a la eficacia y a la fiabilidad de la comunicación al permitir que los comunicadores argumenten sus afirmaciones y que los destinatarios los evalúen. De este modo, aumenta tanto en cantidad como en la calidad epistémica la información que los seres humanos son capaces de compartir.

Consideramos que la evolución del razonamiento está ligada a la de la comunicación humana. El razonamiento, hemos afirmado, permite a los comunicadores producir argumentos para convencer destinatarios que no aceptaban lo dicho por mera confianza; eso permite a los destinatarios evaluar la solidez de los argumentos y aceptar información valiosa que, de otro modo, sería sospechosa. Así, gracias al razonamiento, la comunicación humana se torna más fiable y potente. A partir de la hipótesis de que la función principal del razonamiento es argumentativa, se derivaron una serie de predicciones que, hemos tratado de mostrar, son confirmadas por pruebas existentes. Es cierto que la mayoría de estas predicciones pueden ser derivadas de otras teorías. Podríamos establecer, sin embargo, que la hipótesis argumentativa proporciona una explicación más fundamentada [more principled explanation] de la evidencia empírica (en el caso del sesgo de confirmación, por ejemplo). En nuestra discusión acerca del razonamiento motivado y de la elección basada en la razón, no sólo convergen nuestras predicciones con las teorías existentes, sino que además hemos tomado prestadas muchas de ellas. Incluso en estos casos podríamos, no obstante, argumentar que nuestro enfoque tiene la ventaja distintiva de proporcionar respuestas claras a las preguntas sobre el por qué: ¿Por qué los seres humanos tienen un sesgo de confirmación? ¿Por qué se involucran en el razonamiento motivado? ¿Por qué basan sus decisiones en la disponibilidad de razones justificativas? Por otra parte, la teoría argumentativa del razonamiento ofrece una perspectiva integradora única: explica amplias franjas de la literatura psicológica dentro de un solo marco general.

Algunas de las pruebas aquí examinadas demuestran no sólo que el razonamiento no nos provee creencias y decisiones racionales de forma fiable, sino también que, en una variedad de casos, incluso puede ser perjudicial para la racionalidad. El razonamiento puede conducir a malos resultados no porque los seres humanos sean malos sino porque buscan sistemáticamente argumentos para justificar sus creencias o sus acciones. La teoría argumentativa, sin embargo, pone este tipo de manifestaciones "irracionales" en una perspectiva novedosa. El razonamiento humano no es un mecanismo general profundamente defectuoso; por el contrario, es un dispositivo

especializado, notablemente eficiente, adaptado a un cierto tipo de interacción social y cognitiva en la que sobresale.

Incluso desde un punto de vista estrictamente epistémico, la teoría argumentativa del razonamiento no pinta un cuadro completamente desesperanzador. Sostiene que existe una asimetría entre la *producción* de argumentos (que implica un sesgo intrínseco en favor de las opiniones o decisiones del argumentador, sean sólidas o no) y la *evaluación* de argumentos (que tiene por objeto distinguir los buenos argumentos de los malos y, de este modo, la información genuina de la errónea). Esta asimetría se encuentra a menudo oscurecida en un debate (o en una situación en la que se puede anticipar que habrá un debate al respecto). En realidad, las personas que tienen una opinión que defender no evalúan los argumentos de sus interlocutores en busca de información verdadera, sino que los consideran desde el principio como contra-argumentos a rebatir. Sin embargo, como lo demuestra la evidencia revisada en la sección 2, las personas son buenas en la evaluación de argumentos y son capaces de hacerlo de manera imparcial, siempre que no se hallen mediados por un interés personal particular. En experimentos de razonamiento grupal, donde los participantes comparten un interés en el descubrimiento de la respuesta correcta, se ha demostrado que *la verdad triunfa* (Laughlin & Ellis 1986; Moshman y Geil 1998). Si bien los participantes en tareas colectivas experimentales suelen producir argumentos en favor de una variedad de hipótesis, de las cuales la mayoría o incluso la totalidad es falsa, coinciden al reconocer argumentos sólidos. Dado que estas tareas tienen una solución válida demostrable, la verdad, en efecto, triunfa. Si consideramos problemas que no tienen una solución demostrable, debemos por lo menos esperar que ganen los buenos argumentos, aunque la verdad no termine triunfando (y, en la sección 2, hemos revisado evidencia que muestra que verdaderamente éste es el caso). Esto puede parecer trivial, pero no lo es ya que demuestra que, contrariamente a las evaluaciones sombrías comunes de las habilidades de razonamiento humanas, la gente es muy capaz de razonar de manera imparcial, al menos cuando se están evaluando argumentos en lugar de producirlos y buscan la verdad en lugar de intentar ganar un debate.

Estas sólidas evaluaciones en situaciones similares, ¿no podrían favorecer del mismo modo a la producción de argumentos? Nótese, primero, que las situaciones en donde un interés compartido en la verdad lleva a los participantes de una tarea grupal a evaluar los argumentos de manera correcta, no son suficientes para llevarlos a producir buenos argumentos. En estas tareas grupales, los participantes proponen al grupo las mismas respuestas que pensaron individualmente. El éxito grupal se debe, principalmente, al filtrado de una variedad de soluciones, conseguidas a través de la evaluación. Cuando las diferentes respuestas que se proponen inicialmente resultan en su totalidad incorrectas, entonces todas pueden ser rechazadas, y se proponen hipótesis total o parcialmente nuevas, las cuales son filtradas a su vez, explicando así cómo pueden desempeñarse mejor los grupos que cualquiera de sus miembros individuales.

Los individuos pensando por su propia cuenta sin beneficiarse de las opiniones de los demás sólo pueden evaluar sus propias hipótesis, pero, al hacerlo, son tanto juez como parte o, mejor dicho, juez y abogado, actitud que no resulta óptima para alcanzar la verdad. ¿No sería posible, en principio, para un individuo el decidirse a generar una variedad de hipótesis en respuesta a alguna pregunta y luego evaluarlas una por una, a la manera de Sherlock Holmes? Lo que hace que Sherlock Holmes sea un personaje tan fascinante son precisamente los cambios rotundos en su pensamiento, operando en un mundo manejado por Conan Doyle, donde lo que deberían ser problemas inductivos tienen, en cambio, soluciones deductivas. Yendo a la realidad, los individuos pueden llegar a desarrollar alguna habilidad limitada para distanciarse ellos mismos de su propia opinión, para considerar alternativas y desde ese lugar ser más objetivos. Presumiblemente esto es lo que hace más o menos el 10% de las personas que pasan la tarea de selección estándar de Wason. Pero es, de hecho, una habilidad adquirida e involucra el ejercicio de un control imperfecto sobre una disposición natural que inclina espontáneamente en una dirección diferente.

Aquí, uno podría estar tentado de señalar que el razonamiento es el responsable de algunos de los grandes logros del pensamiento humano en los dominios epistémicos y morales. Esto es innegablemente verdadero, pero los logros involucrados son colectivos

y resultan de interacciones hechas a lo largo de muchas generaciones (sobre la importancia de las interacciones sociales para la creatividad, incluyendo la creatividad científica, ver Csikszentmihalyi & Sawyer 1006; Dunbar 1997; John-Steiner 2000; Okada & Simon 1997). Toda la empresa científica ha sido estructurada siempre en grupos, desde la Academia Nacional de los Linces hasta el Colisionador de Hadrones. En el dominio moral, los logros como la abolición de la esclavitud son el resultado de intensas discusiones públicas. Hemos señalado que, en condiciones grupales, los sesgos del razonamiento pueden convertirse en una fuerza positiva y contribuir a una buena división de la labor cognitiva. De todas maneras, para destacarse en tales grupos, puede que sea necesario anticipar cómo pueden llegar a ser evaluados los argumentos propios por los demás y ajustarlos adecuadamente. Mostrarlo podría ser una valiosa habilidad adquirida culturalmente, como en las *disputationes* medievales (ver Novaes 2005). Al anticipar objeciones, uno podría incluso ser capaz de reconocer las fallas en las propias hipótesis y revisarlas. Hemos sugerido que esto depende de una habilidad adquirida dolorosamente para llevar a cabo algún control limitado sobre los sesgos propios. Incluso entre los científicos, esta habilidad puede ser poco común, pero aquellos que la poseen podrían llegar a tener una gran influencia en el desarrollo de ideas científicas. Sería un error, sin embargo, tratar sus contribuciones casi anormales y altamente visibles, como ejemplos paradigmáticos de razonamiento humano. En muchas discusiones, en vez de buscar fallas en nuestros propios argumentos, es más fácil dejar que otra persona lo haga y encargarnos sólo de ajustar nuestros argumentos si resulta necesario.

En general uno debería ser cauteloso al usar los importantes logros del razonamiento como prueba de su eficiencia general, dado que sus fallas son mucho menos visibles (ver Ormerod 2006; Taleb 2007). Los éxitos epistémicos pueden depender, hasta un grado significativo, de lo que los filósofos han llamado *la suerte epistémica* (Pitchard 2005); esto es, los factores de suerte que llegan a ponernos en el camino correcto. Cuando sucede que uno está en el camino correcto y "más en lo cierto" de lo que uno hubiera imaginado al comienzo, algunos de los efectos distorsivos del

razonamiento motivado y de la polarización pueden convertirse en bendiciones. Por ejemplo, el razonamiento motivado puede haber llevado a Darwin a enfocarse obsesivamente en la idea de la selección natural y explorar todos los argumentos posibles en favor de esa idea, y sus consecuencias. Pero, por cada Darwin, ¿cuántos Paleys hay?

Para concluir, señalamos que la teoría argumentativa del razonamiento debería congeniar con los que disfrutamos de interminables debates sobre ideas – pero esto, claro, no es un argumento que confirme (o refute) la teoría.

Agradecimientos de los autores

Estamos agradecidos con Paul Bloom, Ruth Byrne, Peter Carruthers, Nick Chatter, Jon Haidt, Ira Noveck, Guy Politzer, Jean-Baptiste Van der Henst, Deirdre Wilson, y con los cuatro revisores anónimos por sus útiles sugerencias y críticas a las versiones previas de este artículo. Nuestro trabajo ha sido posible gracias al apoyo de la beca doctoral de la DGA (París) otorgada por la CSMN (Oslo) a Hugo Mercier.