
AUTOCONOCIMIENTO Y LA PSICOLOGÍA EXPERIMENTAL

MARTÍN FRANCISCO FRICKE

El conocimiento que, al parecer, tenemos de nuestros propios estados mentales es especial. Podemos saber qué es lo que pensamos, sentimos o aparentemente percibimos de una manera más inmediata que cuando se trata de los estados mentales de otras personas. En el caso de otros, tenemos que observar su comportamiento, escuchar sus emisiones y sacar nuestras conclusiones a partir de estos datos. Pero en el caso de nosotros mismos parece que no es necesario que nos observemos e infiramos a partir de las propias emisiones qué es lo que pensamos o sentimos. Además, parece que es muy poco probable, e quizá imposible, que nos equivoquemos en este tipo de conocimiento. Cuando sinceramente afirmo “Creo que X ganará las elecciones”, fácilmente puedo equivocarme con respecto a las elecciones. Quizá otra persona ganará las elecciones. Pero difícilmente me equivoco acerca del hecho de que *yo creo* que X ganará las elecciones. Estos dos aspectos del autoconocimiento, el acceso comparativamente directo y la poca probabilidad de error, caracterizan el fenómeno de la autoridad de la primera persona. ¿Cómo se explica este fenómeno? ¿Cómo es posible que tengamos tal autoridad?

Desde Descartes, muchos filósofos han tratado de contestar estas preguntas y de maneras muy diversas. Descartes mismo, por ejemplo, sugiere que tenemos una facultad introspectiva que nos informa con certeza infalible del contenido de nuestras mentes. En la filosofía contemporánea algunos mantienen la idea de una facultad introspectiva, pero la identifican con un proceso falible en el cual una parte del cerebro escanea a otra (Armstrong 1993). Davidson (1984) argumenta que la autoridad de la primera persona se basa en el conocimiento que tenemos del significado de nuestras palabras y que este conocimiento necesariamente es autoritativo —debido a la naturaleza del proceso de interpretación por medio del cual el significado se da. Otros filósofos contemporáneos sugieren que el autoconocimiento es autoritativo por una relación constitutiva entre las creencias sobre nuestros estados mentales y estos estados mentales mismos (Wright 1989, Heal 2003, Stoneham 1998). Todavía otros piensan que

Instituto de Investigaciones Filosóficas y Centro Peninsular en Humanidades y Ciencias Sociales, Universidad Nacional Autónoma de México. / martin-fricke@yahoo.co.uk

nuestra racionalidad necesariamente requiere que tengamos un conocimiento especial de la propia mente (Shoemaker 1988, Burge 1996). También hay filósofos que defienden la idea de un acceso especial a la propia mente, pero no basada en una capacidad introspectiva sino en una que se podría describir como extrospectiva: dirigiendo mi atención a hechos en el mundo externo, conozco mis creencias sobre el mismo (Evans 1982, Byrne 2005). Finalmente, también hay filósofos escépticos. Ellos no aceptan la idea de una asimetría fundamental entre el conocimiento de mis propios estados mentales y el de los estados mentales de otros. Según ellos, la manera de conocer los estados mentales es la misma en ambos casos y si conozco mejor mi propia mente eso sólo se debe a la mayor cantidad de datos que tengo sobre mí mismo (Ryle 1949, Carruthers 2010).

¿Cómo podemos evaluar estas teorías y cómo decidir cuál o cuáles de ellas son correctas? Un elemento en tal evaluación debe ser la coherencia de la teoría. Quizá evaluar la coherencia de teorías es la competencia clave de los filósofos. Es una tarea que depende muy poco de un acceso a datos empíricos, tal vez es una tarea que opera en el ámbito de lo *a priori*.

Pero la coherencia no es todo en la evaluación de teorías. Puede ser que varias teorías sean igualmente coherentes internamente, pero no mutuamente. En este caso, por lo menos una de las teorías es falsa y la evaluación debe incluir, además de la coherencia, otros criterios de verdad. Podemos decir que la evaluación debe decidir cuáles teorías son más plausibles que las otras. Aquí sí necesitamos datos empíricos para diferenciar entre las teorías. (Teóricos coherentistas de la verdad dirían que incluimos más proposiciones en los conjuntos cuya coherencia evaluamos.)

¿De dónde tomar estos datos empíricos? Muchos datos empíricos son lugares comunes. Sabemos que somos seres humanos, que hay 'objetos secos de tamaño mediano' tales como sillas y mesas en el mundo ("*dry middle-sized goods*" como John Austin supuestamente los llamaba), que nos comunicamos con otras personas, incluso sabemos, muchas veces, qué es lo que piensan, sienten y quieren otras personas. Quizá tales datos empíricos generales sean suficiente apoyo empírico para la mayoría de las teorías filosóficas. Toda la filosofía crítica de Kant, aunque hace afirmaciones sobre algunas ciencias particulares como la física, en sus argumentos más fundamentales, parece descansar sólo en tales hechos generales sobre el mundo. Ciertamente, la teoría interpretacionista del lenguaje (y de la autoridad de la primera persona) que ofrece Davidson sólo se apoya en tales hechos generales. La teoría de Wright, inspirada en las *Investigaciones filosóficas* de Wittgenstein, empíricamente sólo está basada en consideraciones muy generales sobre nuestro uso del lenguaje, el cual toma como evidente para cualquier filósofo que lo contempla sinceramente. En otras palabras, estas teorías no se basan en hechos empíricos que sólo pueden ser descubiertos por expertos que se especializan en investigar esta parte

del mundo empírico. Más bien, se supone que los hechos empíricos en cuestión sean tan generales, sencillos y evidentes a nuestra atención que no necesitan de especialistas para ser descubiertos.

Sin embargo, también hay teorías filosóficas que versan sobre detalles más esotéricos del mundo empírico, detalles que no son sencillos y evidentes a nuestra atención. Por ejemplo, en la ética médica se discute si en ciertas situaciones es permitido o quizá incluso deseable terminar la vida de alguien que ha perdido las funciones de su cerebro y sólo se mantiene con vida gracias a aparatos médicos. Esta discusión ética probablemente no puede conducirse fructíferamente si no se conocen hechos empíricos relevantes, sobre la función del cerebro, los aparatos médicos etcétera, de los cuales sólo se puede saber a través de expertos médicos. Igualmente, no tiene mucho sentido tratar de desarrollar una teoría sobre algún aspecto de la teoría de la evolución tal como la selección de grupos o sobre la física cuántica o sobre cualquier otro tema que es susceptible para ser examinado por alguna ciencia particular, sin consultarse con especialistas que investigan el tema en cuestión.

En el caso que me concierne en mis investigaciones, las ciencias particulares que podrían contribuir hechos relevantes son la psicología experimental, las ciencias del cerebro y quizá otras ramas de las ciencias cognitivas. Peter Carruthers (2010) es un autor que recientemente ha tratado de utilizar resultados de estas disciplinas para justificar algunas afirmaciones interesantes sobre el autoconocimiento. Mencionaré tres de los resultados psicológicos que Carruthers utiliza.

(1) Pacientes en los cuales la conexión de nervios entre los dos hemisferios del cerebro ha sido cortada ("*split brain*"). Carruthers (2010: 84-87) se refiere a un experimento reportado por Gazzaniga¹ (1995). Se presentaron estímulos diferentes a los dos hemisferios. Por ejemplo, una pata de pollo en el lado derecho del campo visual (accesible sólo al hemisferio izquierdo del cerebro) y una escena nevada en el lado izquierdo (accesible sólo al hemisferio derecho). Luego se pidió al paciente elegir imágenes asociadas a lo que había visto desde un conjunto de fotos (visualmente accesibles a los dos hemisferios). Con su mano izquierda, controlada por el hemisferio izquierdo, el sujeto eligió la foto de un pollo, con la otra mano, controlada por el hemisferio derecho, una pala. Evidentemente, el hemisferio izquierdo asoció correctamente el pollo con la pata de pollo y el hemisferio derecho correctamente la pala con la escena nevada. (El contexto es el norte de los Estados Unidos, donde regularmente se da la necesidad de limpiar la nieve con palas.) Sin embargo, cuando se preguntó al paciente por qué había elegido esos objetos, explicó: "El pollo va con la pata de pollo, y se necesita una pala para limpiar el establo de los pollos". Esta explicación verbal fue producida por el hemisferio izquierdo, el cual no tenía acceso a la información que se presentó al hemisferio derecho. El paciente *confabuló*

una explicación sobre sus propios juicios. No reportó una falta de datos, ni profesó ignorancia. Carruthers infiere que, de hecho, no tenemos un acceso directo introspectivo a los propios juicios. Más bien, *interpretamos* (inconscientemente) nuestras acciones y el contexto para inferir qué es lo que juzgamos o decidimos —de la misma manera en que averiguamos qué es lo que juzgan o deciden otras personas. La idea de un acceso privilegiado a la propia mente es, entonces, ficticia.

(2) En el segundo caso (Carruthers 2010: 89s) un sujeto (sano) recibe bajo hipnosis la instrucción de tomar un libro de la mesa y de ponerlo en el librero. Al despertar de la hipnosis el sujeto cumple con la orden. Frecuentemente, el sujeto no recuerda haber recibido una orden ni haber sido hipnotizado. En tales casos, cuando se le pregunta al sujeto por qué tomó el libro y lo puso en el librero éste a menudo confabula una explicación, diciendo que no le gusta el desorden y que por eso levantó el libro (Wegner 2002). Nuevamente, el sujeto no reporta ignorancia de sus propios motivos sino parece interpretarse a sí mismo, de la misma manera en que una tercera persona probablemente lo haría si lo observara y no supiera de la hipnosis. De nuevo, Carruthers concluye que ordinariamente no introspeccionamos nuestras decisiones directamente. Sólo los conocemos a través de una interpretación que procede inconscientemente de la misma manera que cuando averiguamos las decisiones de otras personas.

(3) El tercer experimento que quiero mencionar es de la psicología social (Carruthers 2010: 95-97). Unas personas son manipuladas de tal forma que escriben un ensayo defendiendo una tesis con la cual inicialmente no están de acuerdo y reciben una compensación financiera miserable. Otras personas hacen lo mismo pero reciben un pago decente. En tal situación, las personas en el primer grupo más frecuentemente llegan aceptar que “tienen simpatías” con las tesis que defendieron en sus ensayos (Cohen 1962, Lindner, et al. 1967). Carruthers sugiere que eso se debe a una interpretación que podría ser explicitada en el siguiente razonamiento: “Ya que invertí tanto tiempo en escribir un ensayo por tan poquito dinero y ya que lo hice de mi propia elección, debe ser que la posición que defendí me pareció correcta y merecedora de ser defendida. Por tanto, debe ser que la creo” (Carruthers 2010: 95).

Los tres experimentos —y varios otros descritos por este autor— señalan lo mismo, según Carruthers no introspeccionamos los propios juicios y decisiones. No tenemos un acceso inmediato y exclusivo de la primera persona a los propios juicios y decisiones. Más bien, como en el caso de la atribución de juicios y decisiones a otras personas, realizamos una *interpretación* de nuestro propio comportamiento. Como demuestran los experimentos, tal interpretación puede fallar. Si no es tan susceptible a la posibilidad del error como atribuciones a otras personas eso se debe

únicamente a la mayor cantidad de datos que cada uno tiene sobre sí mismo.

¿Es la interpretación que Carruthers ofrece de los citados experimentos psicológicos correcta? Quizá no. Pero me parece que difícilmente se puede negar que una teoría de la autoridad de la primera persona debería discutirlos para ser completa. Por lo menos para la filosofía del autoconocimiento, entonces, es fructífero y necesario tomar en cuenta los resultados de otras disciplinas científicas².

¿Es también la filosofía útil para estas otras disciplinas? Me parece que numerosos ejemplos muestran que sí, pero este argumento debería ser elaborado en otra contribución a este foro.

NOTAS

- 1 Me refiero al texto de Gazzaniga (1995) así como, en lo que sigue, a los de Wegner (2002), Cohen 1962 y Lindner et al. (1967) a través del artículo de Carruthers (2010: 84ss.).
- 2 Las disciplinas "científicas" incluyen, según mi parecer, las humanidades, conocidas en países de habla alemana como "ciencias del espíritu" (*Geisteswissenschaften*).

BIBLIOGRAFÍA

- Armstrong, D. M. (1993), *A Materialist Theory of the Mind*, rev. edition, London: Routledge.
- Burge T. (1996), "Our entitlement to self-knowledge," *Proceedings of the Aristotelian Society* 96: 91-116.
- Byrne, A. (2005), "Introspection," *Philosophical Topics* 33: 79-104.
- Carruthers, P. (2010), "Introspection: divided and partly eliminated", *Philosophy and Phenomenological Research* 80: 76-111.
- Cohen, A. (1962). "An experiment on small rewards for discrepant compliance and attitude change," in J. Brehm y A. Cohen (eds.), *Explorations in Cognitive Dissonance*, NY: Wiley.
- Davidson, D. (1984), "First person authority", *Dialectica* 38: 101-111 (re-impresión en *Subjective, Intersubjective, Objective*, Oxford: Clarendon, 2001).
- Evans, G. (1982), *The Varieties of Reference*, Oxford: Clarendon.
- Gazzaniga, M. (1995). "Consciousness and the cerebral hemispheres," in M. Gazzaniga (ed.), *The Cognitive Neurosciences*, Boston: MIT Press.
- Heal, J. (2003), "On first-person authority," in J. Heal, *Mind, Reason and Imagination*, Cambridge: Cambridge University Press, pp. 273-288.
- Linder, D., Cooper, J. y Jones, E. (1967) "Decision freedom as a determinant of the role of incentive magnitude in attitude change," *Journal of Personality and Social Psychology* 6: 245-254.
- Ryle, G. (1949), *The Concept of Mind*, London: Hutchinson.
- Shoemaker, S. (1988), "On knowing one's own mind," *Philosophical Perspectives* 2: 183-209.
- Stoneham, T. (1998), "On believing that I am thinking," *Proceedings of the Aristotelian Society* 98: 125-144.
- Wegner, D. (2002), *The Illusion of Conscious Will*, Boston: MIT Press.
- Wright, C. (1989), "Wittgenstein's later philosophy of mind. Sensation, privacy and intention," *Journal of Philosophy* 86: 622-634.