



Reproducibilidad y repetitividad experimental entre filósofos y científicos

Marisa Velasco*

Introducción

Hace casi una década que escuchamos hablar cada vez más a menudo de “crisis de reproducibilidad” en las ciencias experimentales. Las más prestigiosas revistas científicas dedican casi un número al año, muchos editoriales, cartas de lectores y encuestas e investigaciones entre sus lectores sobre el problema de la crisis de reproducibilidad. Muchas publicaciones han cambiado sus políticas editoriales y los mecanismos de aceptación y evaluación por estas razones. Desde aproximadamente 2010 la literatura en revistas científicas sobre el problema es muy abundante, y a su vez en ciertas áreas disciplinares se han creado organizaciones para enfrentar el problema.

Por otra parte, los filósofos de la ciencia han discutido profusamente el problema de la replicabilidad y la reproducibilidad experimental, especialmente desde la década del 80 del siglo pasado.

Sin embargo, es posible notar que las reflexiones de científicos y filósofos van en direcciones en parte divergentes y que, al menos, existe una tensión en el modo en que caracterizan y el rol que dan a la reproducibilidad y replicabilidad experimental.

En la primera parte de este trabajo reconstruiré de un modo muy breve los conceptos filosóficos de replicabilidad, repetitividad y reproducibilidad experimental, así como los problemas asociados con esos conceptos. La literatura es abundante y las problemáticas discutidas, así como las posiciones asumidas, por los filósofos son muy diversas. Por razones de espacio, y porque esta literatura es más conocida en el ámbito filosófico, sólo presentaré algunas distinciones y problemas asociados.

En una segunda parte, intentaré caracterizar y sistematizar los múltiples problemas que se esconden detrás del concepto de “crisis de reproducibilidad” en la literatura científica. Esta presentación intentará mostrar

* Universidad Nacional de Córdoba, Facultad de Filosofía y Humanidades.
marisa@ffyh.unc.edu.ar



que algunos de esos problemas están muy lejos de las reflexiones filosóficas y otros, aunque están más cerca, han recibido un tratamiento completamente diferente al que recibieron en ámbitos filosóficos.

El problema entre filósofos de la ciencia

En gran parte de los textos clásicos de filosofía de la ciencia se puede encontrar reflexiones sobre el tema, generalmente tratando de elucidar por qué la replicabilidad es la marca del método científico. Por ejemplo,

Sólo cuando se da la recurrencia de ciertos acontecimientos de acuerdo con reglas o regularidades —y así sucede con los experimentos repetibles— pueden ser contrastadas nuestras observaciones por cualquiera (en principio). Ni siquiera tomamos muy en serio nuestras observaciones, ni las aceptamos como científicas, hasta que las hemos repetido y contrastado. Sólo merced a tales repeticiones podemos convencernos de que no nos encontramos con una mera «coincidencia» aislada, sino con acontecimientos que, debido a su regularidad y reproductibilidad, son, en principio, contrastables intersubjetivamente. . . . Puede definirse el efecto físico científicamente significativo como aquél que cualquiera puede reproducir con regularidad sin más que llevar a cabo el experimento apropiado del modo prescrito. (Popper, 1962, p. 44)

En Collins (1975) podemos encontrar la primera versión de su tesis del “regreso del experimentador”. Esta idea fue desarrollada luego con detalle en *Changing order* (1985) y posteriormente en diversos artículos. En forma breve, y en las propias palabras del autor, la tesis sostiene que:

. . . el *regreso del experimentador* . . . es una paradoja que se produce para aquellos que usan la replicación como un test de la verdad de las afirmaciones del conocimiento científico. El problema es que, puesto que la experimentación es un problema de habilidades prácticas, nunca puede estar claro si un segundo experimento que ha sido realizado suficientemente bien cuenta para chequear los resultados del primero. Algún test adicional es necesario para testear la calidad del experimento —y así sucesivamente. (Collins, 1985, p. 2)

¡Pero no sabremos si hemos construido un buen detector hasta que lo usemos y obtengamos el resultado correcto! (Collins, 1985, p. 84)

La primera es la versión experimental, la segunda es la versión instrumental de la tesis. Las repercusiones de las ideas de Collins han sido muy

variadas, así como las interpretaciones que se han dado de esta tesis. Hasta hace unos pocos años la más común era que Collins no estaba sosteniendo un argumento epistemológico, sino uno sociológico, porque sólo puede romperse el regreso infinito a través de factores no epistémicos. La idea de conocimiento implícito, a la que recurrió el propio Collins, no jugaría en estas interpretaciones un papel epistémico. Estas discrepancias, en el marco de las disputas con las perspectivas sociales de la ciencia, llevaron, por ejemplo, a encendidas discusiones entre Allan Franklin y Harry Collins. Sin embargo, hoy es posible encontrar un escrito común entre Franklin y Collins (2016) hablando exactamente de este problema en el que reconocen más acuerdos que discrepancias.

A partir de las ideas iniciales de Collins, y especialmente después del nacimiento de lo que dio en llamarse epistemología de la experimentación, es posible encontrar muchas reflexiones filosóficas sobre replicabilidad y reproducibilidad experimental, aunque no necesariamente en continuidad con la tesis de Collins.¹

La literatura filosófica no ha sido demasiado uniforme en el uso terminológico, aunque sí en las distinciones conceptuales por lo que utilizaré los términos con el siguiente significado:

- **Repetibilidad** del experimento en sus procedimientos y condiciones materiales.
- **Reproducibilidad** del resultado de un experimento a través de diferentes procedimientos materiales para obtener el mismo resultado experimental.
- **Replicabilidad** experimental incluye, como término genérico, la repetibilidad y la reproducibilidad.²

Esta literatura ha abordado como tópicos de interés filosófico los siguientes problemas:

¹ Cf., por ejemplo, entre muchos otros trabajos, Radder (1992, 2003), Steinle (2002, 2016), Bogen (2001), Franklin (1994, 1997), Leonelli (2018), Norton (2015), Schickore (2010, 2011), Feest (2016, 2019).

² El vocabulario no sólo no es uniforme, sino que además se incorporan otras distinciones como, por ejemplo, reproducción directa y reproducción conceptual, pero que no aportan a lo que este trabajo quiere mostrar. La terminología adoptada aquí sigue casi textualmente la adoptada por Radder, aunque no necesariamente se siguen las ideas de este autor.

- a) La posibilidad misma de que un experimento sea repetible. Es a partir de tratar de elucidar qué significa repetir un experimento que los conceptos anteriores (repetibilidad, reproducibilidad y replicabilidad) aparecen como distinciones necesarias.
- b) Sólo los experimentos repetibles/reproducibles (repetidos/reproducidos) constituyen evidencia científica. Poner en cuestión esto implica, por un lado, distinguir los experimentos repetibles de los experimentos repetidos, pero, por otro lado, también obliga a repensar el propio concepto de evidencia experimental y su rol dentro del conocimiento científico.
- c) La replicabilidad experimental es la marca del método científico, al menos desde el siglo XVII. Diversos trabajos han mostrado el carácter histórico de la noción de replicabilidad, esto es, han mostrado diferentes modos de entender la replicabilidad en distintas disciplinas durante los siglos XVII, XVIII y XIX. Pero estos trabajos también han mostrado que la no-replicabilidad experimental ha tenido diferentes roles en diferentes períodos históricos.
- d) El problema de la confiabilidad de los experimentos únicos. Esto es, las prácticas científicas muestran experimentos que han sido diseñados para no ser repetidos (por diversas razones prácticas y teóricas) y sin embargo sus resultados gozan de una gran credibilidad.
- e) El control instrumental es un modo alternativo de replicabilidad. Esto es, gran parte de la confianza en los resultados experimentales no proviene de su replicabilidad, sino de que podamos tener un control y calibración instrumental como base de esa confianza.

Muchos de los tópicos de esta lista, que está muy lejos de ser exhaustiva, están profundamente relacionados y en algunas ocasiones han sido la base del desarrollo de programas de investigación en filosofía de la ciencia. Mi intención con esta enumeración es hacer patente que estos son problemas acerca de los cuales han pensado muchos filósofos de la ciencia de modo explícito desde por lo menos la década del 1980.

Aunque los modos de abordar estos problemas y las soluciones propuestas por los filósofos de la ciencia no son uniformes, como es habitual en esta disciplina, creo que casi todos ellos acordarían con la siguiente cita:

Distintos casos históricos muestran una sorprendente variedad de actitudes hacia la replicabilidad. Esto deja en claro que la replicabilidad no fue ni es el único medio para aceptar los resultados experimentales. . . . La replicabilidad puede ser una condición suficiente pero no es una condición necesaria para la aceptación de los resultados experimentales. (Steinle, 2016, pp. 57-58)

Reproducibilidad entre científicos: “crisis de reproducibilidad”

Entre el 2010 y 2015 se realizaron una serie de estudios sobre problemas de reproducibilidad de resultados publicados en prestigiosas revistas científicas³. Los primeros estudios afectan de modo especial a investigaciones en psicología e investigaciones pre-clínicas de cáncer y muestran tasas muy altas de problemas de replicabilidad experimental. Esta discusión en el ámbito científico recibe comúnmente el nombre de “crisis de reproducibilidad”. Aunque el vocabulario también es variable en este ámbito en general se refieren alternativamente a lo que hemos denominado más arriba como repetibilidad, reproducibilidad y replicabilidad, alternativamente. Los problemas, perspectivas y disciplinas son muy amplios y variados en la literatura de científicos e investigadores de lo que se ha denominado *metascience*. Sin embargo, el marco general que comparten podría resumirse en la siguiente cita que introduce la presentación de un simposio de científicos sobre la temática:

La *regla de oro* de la ciencia es la reproducibilidad. Idealmente, las investigaciones son dignas de atención, publicación y cita si investigadores independientes pueden reproducirlas. . . . Pero en gran parte de la literatura los resultados no son reproducibles. (Grant, 2012, el énfasis es mío)

Son muy pocas las ocasiones donde la “regla de oro” se pone en tela de juicio y es por esta razón que la llamo el marco de la discusión y ese marco explica por qué las dificultades en la replicabilidad dan lugar a una crisis.

³ Podrían citarse muchos estudios en esta dirección, así como mucha discusión, en el ámbito de las revistas científicas más prestigiosas, sirvan a modo de ejemplo (D. Baker, Lidster, Sottomayor, & Amor, 2012; M. Baker, 2015; Begley, 2013; Begley & Ellis, 2012; Open Science Collaboration, 2015; Schooler, 2014). Tanto es así que casi todas ellas tienen anualmente a partir de esa fecha algún editorial y alguna sección especial dedicada al tema. *Nature* realizó una encuesta entre sus lectores y autores en 2017 cuyos resultados y análisis pueden encontrarse en <https://media.nature.com/original/magazine-assets/d41586-018-04590-7/15675426>. Las publicaciones en esa misma revista sobre problemas de reproducibilidad entre 2012-2018 puede encontrarse en <https://www.nature.com/collections/prbfkwmvzw>

Aunque a lo largo de estos años la literatura científica ha abordado diversos temas en torno a lo que denominaron crisis de reproducibilidad, la siguiente lista muestra los tópicos más recurrentes en la discusión:

- a) Se da por sentado que habitualmente los resultados publicados no son reproducidos, ni repetidos por la comunidad científica. Esto recibe tratamiento de diverso tipo: las revistas importantes no publican normalmente reproducciones ni repeticiones experimentales, sean estas exitosas o no, los científicos jóvenes pierden tiempo de carrera académica si se dedican a este tipo de actividad. Buena parte de la discusión tomó esto como un hecho que lleva a la discusión del punto siguiente.
- b) Las publicaciones de resultados experimentales rara vez incluyen datos suficientes para llevar adelante repeticiones experimentales, y hasta puede resultar complejo llevar adelante reproducciones. Aunque en algunas disciplinas resulte complejo pensar qué significaría repetir un experimento, en general la regla de oro no es puesta en tela de juicio. Solo en trabajos más recientes esta variable ha empezado a tenerse más en cuenta.
- c) El tratamiento estadístico de los datos publicados no siempre resulta transparente en una publicación, por lo que se sugiere que el problema no es el diseño y la ejecución experimental sino la inferencia estadística involucrada en los datos publicados.
- d) Casi como una consecuencia de los datos anteriores (y porque además es un problema para las disciplinas que más impacto recibieron por el análisis de reproducibilidad) a veces parece que el problema dejaría de ser tal si las muestras experimentales tuvieran el tamaño adecuado para las inferencias que se llevan adelante con esos resultados.
- e) Esta literatura parece suponer, y a veces lo enuncia explícitamente, que la crisis de reproducibilidad no está uniformemente repartida a través de las disciplinas, las más afectadas parecen ser la psicología, los tratamientos clínicos (especialmente los de cáncer) y las investigaciones ligadas a neurociencias.⁴

⁴ Resulta llamativo que en la discusión al interior de algunas de estas disciplinas parece suponerse, aunque no se diga de modo explícito, que hay disciplinas o áreas disciplinares que resultan casi inmunes a los problemas de replicabilidad. Más adelante se listan algunas estrategias para salir de la crisis de reproducibilidad, muchas de las estrategias hacen patente que habría disciplinas que constituyen modelos a seguir.

- f) Finalmente, aunque en términos epistemológicos puede resultar el aspecto menos relevante, se discute si la crisis de reproducibilidad es una forma de fraude dentro de la ciencia.⁵

Esta discusión, que aquí ha sido fuertemente resumida y esquematizada, tuvo diversos impactos en los modos de hacer y publicar resultados científicos. La siguiente lista, también de modo resumido y esquemático, muestra las respuestas a los problemas planteados por parte de las institucionales y a través de la modificación de prácticas de investigación:

- a) Cambios en políticas editoriales. Las revistas científicas exigen más datos, generalmente en forma de anexos disponibles al menos para evaluadores, antes de la evaluación para publicación, la mayoría de las revistas han elaborado *check lists* para el envío de trabajos empíricos basadas en problemas detectados por esta discusión, se han incorporado evaluadores especialistas en estadística, que no son necesariamente especialistas disciplinares, entre otras modificaciones.
- b) Movimiento *Open Science*. En buena medida el movimiento de Ciencia Abierta es una respuesta a esta discusión. Se espera que la disponibilidad “completa”⁶ y abierta de los resultados experimentales subsanen en gran parte los problemas que la crisis de reproducibilidad ha dejado al descubierto.
- c) Trazabilidad experimental. Como consecuencia de los puntos anteriores (de ambos, o al menos de uno de ellos) se espera que la trazabilidad experimental sea una fuente de transparencia y confiabilidad en los resultados que la crisis puso en tela de juicio.
- d) “Elaboración y estandarización metodológica” y reproducción parcial. Los menos optimistas respecto a que los puntos anteriores logren evitar los problemas puestos en evidencia por la crisis de reproducibilidad, aunque sin renunciar a la regla de oro, han puesto sus esperanzas en estos dos aspectos. El supuesto es que disciplinas altamente estan-

⁵ En este punto hay un acuerdo bastante unánime que la incidencia del fraude intencional tiene un impacto poco significativo. Lo que hace que la crisis de reproducibilidad tenga consecuencias principalmente epistemológicas y metodológicas.

⁶ No he podido evitar las comillas porque la propia idea de completa, así como la idea de estandarización ideal del punto d) de esta lista, resulta cuanto menos ingenua desde una perspectiva filosófica.

darizadas en sus protocolos experimentales no tienen los problemas antes mencionados, y que si no es posible la replicabilidad al menos se debe aspirar a una reproducción parcial. Estas ideas son producto de discusiones que intentan interpretar o poner en tela de juicio los resultados que presenta esta disciplina llamada *meta-science*, por lo que bien podríamos llamarla *meta-meta-science*.

Consideraciones finales

Los filósofos y los historiadores de la ciencia parecen haber puesto en cuestión lo que hemos llamado la “regla de oro” de la ciencia, al menos en una versión irrestricta. De esta forma la replicabilidad experimental resulta en uno de los múltiples modos en que la evidencia experimental gana en confiabilidad. Los científicos parecen no seguir tampoco esta regla en sus prácticas. Esto genera, al menos en la versión pública de la ciencia, desconfianza en los resultados publicados, así como un conjunto de reflexiones y acciones para salir de lo que han llamado “crisis de reproducibilidad”. En síntesis, la tensión parece profunda entre científicos y filósofos en lo que respecta al rol de la reproducibilidad. Sin embargo, las estrategias sugeridas, y en algunos casos llevadas adelante, por los científicos tienden a acercar posiciones.

Referencias

- Baker, D., Lidster, K., Sottomayor, A., & Amor, S. (2012). Reproducibility: Research-reporting standards fall short. *Nature*, *492*(7427), 41.
- Baker, M. (2015). First results from psychology’s largest reproducibility test. *Nature*. doi: 10.1038/nature.2015.17433
- Begley, C. G. (2013). Six red flags for suspect work. *Nature*, *497*(7450), 433-434. doi: 10.1038/497433a
- Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531.
- Bogen, J. (2001). ‘Two as good as a hundred’: Poorly replicated evidence in some nineteenth-century neuroscientific research. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *32*(3), 491-533. doi: 10.1016/S1369-8486(01)00013-9

- Collins, H. M. (1975). The seven sexes: A study in the sociology of a phenomenon, or the replication of experiments in physics. *Sociology*, 9(2), 205-224. doi: 10.1177/003803857500900202
- Collins, H. M. (1985). *Changing order: Replication and induction in scientific practice*. London; Beverly Hills: Sage Publications.
- Feest, U. (2016). The experimenters' regress reconsidered: Replication, tacit knowledge, and the dynamics of knowledge generation. *Studies in History and Philosophy of Science Part A*, 58, 34-45. doi: 10.1016/j.shpsa.2016.04.003
- Feest, U. (2019). Why replication is overrated. *Philosophy of Science*, 86(5), 895-905. doi: 10.1086/705451
- Franklin, A. (1994). How to avoid the experimenters' regress. *Studies in History and Philosophy of Science Part A*, 25(3), 463-491. doi: 10.1016/0039-3681(94)90062-0
- Franklin, A. (1997). Recycling expertise and instrumental loyalty. *Philosophy of Science*, S42-S52.
- Franklin, A., & Collins, H. (2016). Two kinds of case study and a new agreement. En T. Sauer & R. Scholl (Eds.), *The philosophy of historical case studies* (pp. 95-121). doi: 10.1007/978-3-319-30229-4_6
- Grant, B. (2012). Science's reproducibility problem. *The Scientist Magazine*. Recuperado de <https://www.the-scientist.com/news-opinion/sciences-reproducibility-problem-40031>
- Leonelli, S. (2018). Rethinking reproducibility as a criterion for research quality. En L. Fiorito, S. Scheall, & C. E. Suprinyak (Eds.), *Research in the history of economic thought and methodology* (Vol. 36, pp. 129-146). doi: 10.1108/S0743-41542018000036B009
- Norton, J. D. (2015). Replicability of experiment. *THEORIA. An International Journal for Theory, History and Foundations of Science*, 30(2), 229-248. doi: 10.1387/theoria.12691
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). doi: 10.1126/science.aac4716

- Popper, K. R. (1962). *La lógica de la investigación científica* (V. Sánchez de Zavala, Trad.). Madrid: Tecnos.
- Radder, H. (1992). Experimental reproducibility and the experimenters' regress. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 63–73. Recuperado de <http://www.jstor.org/stable/192744>
- Radder, H. (2003). *The philosophy of scientific experimentation*. Pittsburgh, Pa.: University of Pittsburgh Press.
- Schickore, J. (2010). Trying again and again: Multiple repetitions in early modern reports of experiments on snake bites. *Early Science and Medicine*, 15(6), 567–617. doi: 10.1163/157338210X526629
- Schickore, J. (2011). The significance of re-doing experiments: A contribution to historically informed methodology. *Erkenntnis*, 75(3), 325–347. doi: 10.1007/s10670-011-9332-9
- Schooler, J. W. (2014). Metascience could rescue the 'replication crisis'. *Nature News*, 515(7525), 9. doi: 10.1038/515009a
- Steinle, F. (2002). Experiments in history and philosophy of science. *Perspectives on Science*, 10(4), 408–432.
- Steinle, F. (2016). Stability and replication of experimental results: A historical perspective. En H. Atmanspacher & S. Maasen (Eds.), *Reproducibility: Principles, problems, practices, and prospects* (pp. 39–63). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118865064.ch3>