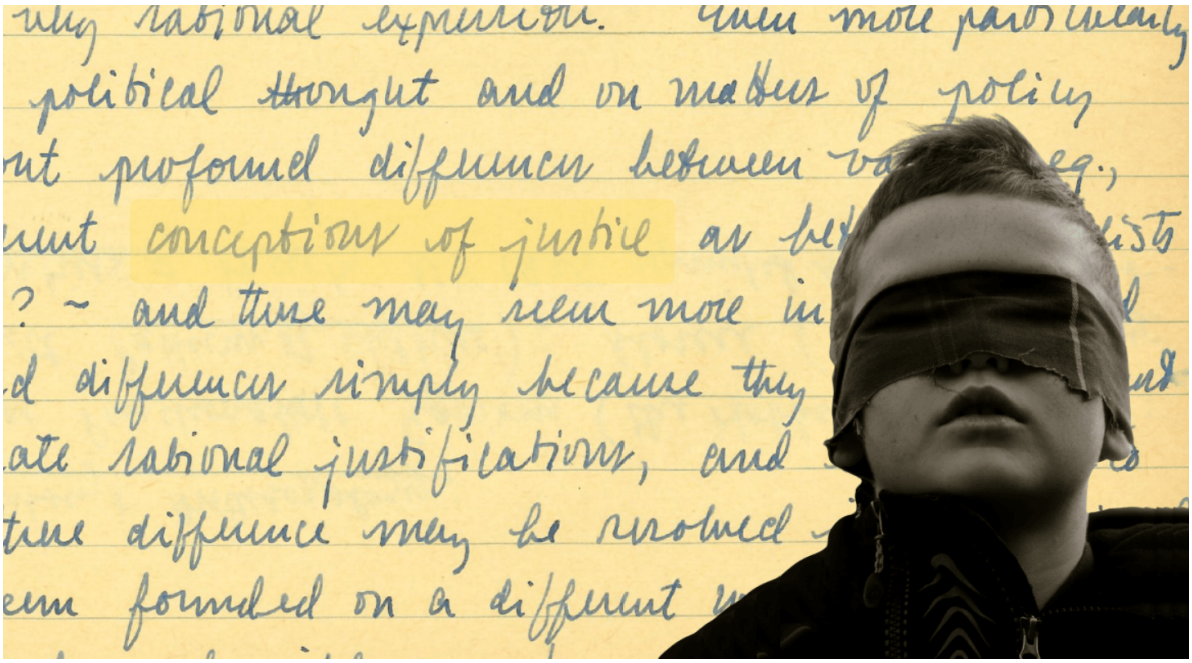
LA POSICIÓN ORIGINAL Y LOS PRINCIPIOS DE JUSTÍCIA

Igualdad y justificación: un replanteamiento de la posición original

Véronique Munoz-Dardé



Manuscrito proveniente del archivo John Rawls Papers (Harvard University Archives). Adaptación gráfica: Maria Oliva Campabadal (CETC).

La posición original tiene una elegancia y un poder superior a la mayoría de imágenes filosóficas. Ha captado la atención de lectores de todo el mundo y de muchas generaciones de estudiantes, además de ser famosa más allá de los círculos filosóficos. Sin embargo, a pesar de su popularidad, la posición original suele distorsionarse y malinterpretarse. En concreto, el método de razonamiento de John Rawls detrás del velo de la ignorancia a menudo se presenta como una forma de alcanzar una conclusión por medio de la teoría de la elección racional. Mi intención en este breve artículo es aclarar el objetivo principal de la posición original y comentar sus características más importantes en contraposición a esta lectura incorrecta dominante.

¿Qué es la posición original?

La posición original es un experimento mental diseñado para expresar de la forma más sencilla y persuasiva las consideraciones que tenemos en nuestras deliberaciones sobre los

principios de justicia social. Rawls sostiene que, juntas, estas consideraciones conducen a sus dos principios de justicia.

La posición original pretende componer una perspectiva equitativa e imparcial sobre la cuestión de justicia. En ella, imaginaremos partes concebidas como personas libres e iguales en desconocimiento de sus características sociales, naturales e históricas como individuos. No conocen su concepción particular de la buena vida, su lugar en la sociedad, situación de clase, estatus social, raza o grupo étnico, sexo, inteligencia, talento natural, etc.; se encuentran tras un v*elo de ignorancia*.

Las partes están detrás del velo de la ignorancia, pero siguen siendo conscientes de los hechos sociales y naturales que afectan a las sociedades humanas. Prevén diferentes conjuntos de principios conforme a los cuales ellas y sus descendientes tendrán que vivir toda su vida. Estos principios se aplicarán a las principales instituciones políticas y sociales que distribuyen conjuntamente los beneficios y las cargas esenciales de la vida social (la *estructura básica* de sociedad). En concreto, las partes tienen en cuenta dos alternativas: una concepción utilitaria de justicia según la cual «una sociedad está correctamente ordenada cuando sus instituciones maximizan el equilibrio neto de satisfacción» [1], y dos principios de Rawls.

En sus dos principios de justicia, Rawls estipula que (a) cada persona tiene derecho al más amplio esquema de libertades básicas compatible con un esquema similar de libertades para todos; y (b) las desigualdades económicas y sociales han de satisfacer dos condiciones: en primer lugar, deben estar adscritos a cargos y posiciones accesibles a todos en condiciones de equitativa igualdad de oportunidades; y, en segundo lugar, tienen que ser para el mayor beneficio de los miembros menos favorecidos de la sociedad (el principio de diferencia) [2].

Rawls afirma que, al reflexionar bajo las limitaciones de la posición original y, por consiguiente, al tener que considerar seriamente qué pasaría si estuvieran entre los miembros menos aventajados de la sociedad, las partes se protegerían a sí mismas contra la pobreza. Asimismo, expone que se adoptarían sus dos principios de justicia en lugar del utilitarismo. Esto significa que las partes rechazarían los principios según los cuales podrían sacrificarse las libertades, las oportunidades o el bienestar de algunos individuos en aras de un bien mayor disfrutado junto con los demás. El menos adinerado tendría una mejor situación si se aplicaran los dos principios de Rawls en lugar de los principios utilitarios: este es el motivo por el que debemos elegirlos.

> La posición original es un experimento mental diseñado para expresar de la forma más sencilla y persuasiva las consideraciones que tenemos en nuestras deliberaciones sobre los principios de justicia social

Desde la publicación de *Teoría de la justicia*, ha habido innumerables argumentos defendiendo que las partes detrás del velo de la ignorancia apoyarían principios diferentes a los que formula Rawls. Estas críticas parecen perder de vista la función que Rawls asigna de manera explícita a la posición original. La tratan como un tipo de experimento en el que descubrimos una conclusión poco evidente a partir de los principios con los que enmarcamos la posición. Enmarcada así, se supone que la posición original *aporta pruebas* a favor de los principios de justicia de Rawls. No obstante, Rawls insiste en que ha enmarcado la posición original con el fin de *ilustrar* los principios y su atractivo racional. Rawls puede estar equivocado sobre a dónde conduce la posición original o quizá sería mejor que eligiera una posición diferente desde la que empezar. Sin embargo, al seguir sus intenciones, debemos esperar una armonía entre la posición original y los principios que recomienda Rawls.

¿Por qué reflexionar sobre la posición original? Disentimientos sobre justicia

¿Qué necesidad hay de reflexionar sobre la posición original? La posición original se ha concebido para poner a prueba si los dos principios pueden ser objeto de un acuerdo alcanzado por todos los miembros de una sociedad caracterizada por profundos desacuerdos sobre qué es justo e injusto.

Escribe Rawls: «La idea intuitiva de la justicia como equidad es considerar los *principios de justicia* como el objeto de un acuerdo original en una situación inicial debidamente definida» [3]. Si aislamos el tipo de consideraciones que pueden justificar derechos sobre justicia, podemos llegar a un estándar público compartido para resolver problemas de las instituciones sociales básicas. Como muchos de sus contemporáneos, Rawls está muy preocupado por los numerosos desacuerdos políticos y morales, que en su época se acentuaron por las controversias sobre los derechos civiles, la discriminación positiva, el aborto, la libertad sexual y la guerra de Vietnam. Por otro lado, también transmite el optimismo relativo de muchos ciudadanos que piensan que es concebible que una sociedad esté «bien ordenada», es decir, que el «sentido público de justicia compartido hace posible que se asocien conjuntamente de manera segura» [4].

Al establecer la reflexión sobre los principios de justicia, Rawls sugiere que es fructífero empezar desde el deseo compartido de una sociedad justa. Afirma que, a pesar de los debates continuos sobre principios políticos, toda la ciudadanía entiende la necesidad de justicia. Los ciudadanos tienen diferentes *concepciones* de justicia y disienten sobre qué es justo e injusto, así como sobre los principios de justicia. No obstante, estas concepciones tienen en común una apreciación de la necesidad de justicia. Por lo tanto, existe un acuerdo subyacente sobre el *concepto* de justicia, definido por esta necesidad:

> «Parece entonces natural pensar que el concepto de la justicia es distinto de las diferentes concepciones de la justicia y que está especificado por el papel que tienen en común estos diferentes

> conjuntos de principios y concepciones. Aquellos que sostienen diferentes concepciones de la justicia pueden […] estar de acuerdo en que las instituciones son justas cuando no se hacen distinciones arbitrarias entre las personas al asignarles derechos y deberes básicos y cuando las reglas determinan un equilibrio debido entre pretensiones competitivas a las ventajas de la vida social».
>
> Rawls, J. (1999) A Theory of Justice. Revised edition. Cambridge Mass.: Harvard University Press, 5 [cursiva añadida].

Por consiguiente, en la búsqueda de un acuerdo sobre los principios de justicia en una sociedad caracterizada por profundos desacuerdos morales, empezamos por los puntos en los que los participantes razonables de una conversación acerca de justicia social pueden estar de acuerdo, al menos en una reflexión. La finalidad de este acuerdo inicial mínimo se centra, en concreto, en las siguientes afirmaciones razonables: (a) no debe haber *distinciones arbitrarias* entre las personas en lo que respecta a derechos y deberes fundamentales, y (b) las ventajas de la vida social, lo que Rawls denomina «los frutos de la cooperación social», deben distribuirse de manera *justa*.

Por lo tanto, la posición original aísla algunas consideraciones razonables compartidas que los ciudadanos *ya apoyan*: cuasitópicos sobre lo que constituye una sociedad justa y razonable, como que sería injusto discriminar a una persona por su sexo, raza u origen étnico.

> La posición original aísla algunas consideraciones razonables compartidas que los ciudadanos ya apoyan sobre lo que constituye una sociedad justa y razonable; por ejemplo, que sería injusto discriminar a una persona por su sexo, raza u origen étnico

El objetivo de Rawls es destacar cómo, al combinarse, estas convicciones morales conducen a principios que otorgan libertades, igualdad ante la ley, condiciones decentes y posibilidades en la vida, así como bases de amor propio, a todos los ciudadanos. Es muy consciente de que a muchas personas les parecerá altamente controvertida la idea de justicia resumida en sus dos principios. Sin embargo, quiere aclarar que las afirmaciones que constituyen premisas ampliamente compartidas entre sus interlocutores *conducen* a estos dos principios. Si está de acuerdo con las afirmaciones anteriores, que apenas son polémicas, entonces está comprometido con sus efectos necesarios. De ahí surge la introducción del velo de la ignorancia y la deliberación sobre la posición original. Rawls espera que el resultado sea aclarar no solo la fuerza de las consideraciones para su propia

concepción de justicia, sino también la naturaleza de los desacuerdos morales con otras posiciones.

La función controvertida de la elección racional

Cualquiera que haya leído *Teoría de la justicia* podrá sospechar que, hasta ahora, he dado muy poca importancia a la función de la elección racional en la posición original y, por supuesto, estará en lo cierto.

Además, en las publicaciones secundarias hay descripciones basadas por completo en la elección racional de la posición original. Desde esta perspectiva, la finalidad de la posición original es servir como una configuración para una elección basada en la teoría de la decisión. Un claro y reciente ejemplo de esta perspectiva es la caracterización de la posición original de Rawls por Lara Buchak [5]. Buchak concibe la posición original como el marco de una teóricaelección basada en la teoría de la decisión relativa a la utilidad esperada y ponderada en función del riesgo de diferentes políticas sociales. Por lo tanto, describe la posición original del siguiente modo:

> «Los individuos tienen en cuenta sus preferencias sobre acuerdos institucionales en la "posición original", en la que las decisiones se toman tras un "velo de ignorancia", donde las personas no conocen con anticipación su "lugar en la sociedad, situación de clase o estatus social, su lugar en la distribución de los recursos y las habilidades naturales, sus objetivos e intereses profundos o su composición psicológica particular". Por consiguiente, los individuos tienen en cuenta sus preferencias sobre una lotería en la que se distribuyen los aspectos sociales y en la que los posibles "estados del mundo" especifican qué lugar ocupará cada uno en la sociedad».
>
> Buchak, Lara, 'Taking risks behind the Veil of Ignorance, Ethics, April 2017, 625 [cursiva añadida].

Lo que resalta la diferencia entre estas estrategias exegéticas es que hay una tensión considerable entre diferentes concepciones de la posición original en la recepción de la teoría de Rawls. Es justo decir que la controversia entre estas concepciones incompatibles es una consecuencia de las propias ambigüedades de Rawls en lo que respecta a la función de la elección racional en la deliberación de los principios en *Teoría*.

La ambigüedad de Rawls, al menos en la versión inicial de *Teoría*, se encuentra entre una explicación de elección puramente racional de la teoría de la política en la que

demostramos a individuos egoístas que la adopción de ciertos principios se hace en su interés prudencial, y una explicación contraria de las teorías políticas en una perspectiva ética más amplia que presenta la posición original como un mero seguimiento de las razones que ya tenemos todos, empezando por un interés compartido en la necesidad de instituciones justas. La última la trata como un experimento de pensamiento en el que descartamos consideraciones que introducirían distorsión o preferencias en el pensamiento sobre la justicia social en lugar de poner énfasis en el egoísmo o en una concepción axiomatizable de lo racional. No obstante, si optamos por esta segunda interpretación, la pregunta obvia es cuál es la función, si la hay, de la elección racional.

¿Cuál es entonces el lugar de la elección racional en la deliberación sobre los principios de justicia? Lo primero que hay que tener en cuenta es que el propio Rawls cambió de opinión tras la publicación de *Teoría*: de considerar que la teoría de la justicia formaba *parte* de la teoría de la elección racional a concebir la elección racional como un mero conjunto de herramientas formales en una concepción de justicia centrada en principios razonables y no en un interés racional. Por consiguiente, en su obra póstuma *Reformulación*, señala:

> «Aquí corrijo que una observación de Teoría de la justicia, donde se dice que la teoría de la justicia forma parte de la teoría de la elección racional. […] Esto es simplemente un error […] Lo que debería haberse dicho es que, para describir a las partes, y su razonamiento, se usa la teoría de la elección (decisión) racional, pero que esta teoría es en sí misma parte de una concepción política de la justicia, una concepción que trata de justificar principios razonables de justicia. En absoluto se pretende derivar esos principios del concepto de la racionalidad como único concepto normativo».
>
> Rawls, J. (2001) Justice as Fairness: A Restatement, edited by Kelly, E. Cambridge, Mass: Harvard University Press, 82 [cursiva añadida].

Lo que Rawls destaca es una cuestión sobre la perspectiva de la persona egoísta que imaginamos deliberando tras el velo de la ignorancia y cómo llega a aceptar los principios de justicia. La cuestión, en palabras de T. M. Scanlon, es si la persona egoísta detrás del velo de la ignorancia está dispuesta a aceptar un principio de justicia «porque considera que *no puede rechazarlo razonablemente sea cual sea la posición que llegue a ocupar* o si, por el contrario, se supone que es aceptable para una persona en cualquier posición social *porque sería la elección racional para una persona egoísta* tras el velo de la ignorancia» [6].

Algunos de los problemas sobre justicia y razonabilidad cuando se leen simplemente como

hechos para la elección racional en la deliberación deben dejar indiferente a un agente: no obligan a nadie a cooperar o no. La posición original puede entenderse como una sugerencia de cómo podemos añadir más limitaciones a la teoría de la elección racional de tal manera que sea racional para el agente jugar a un juego con cierto contenido político.

En varios puntos, parece que Rawls tiene la ambición de demostrar cómo puede deducir una persona principios éticos reales, pertenecientes al campo político, a partir de estas consideraciones racionales tan amplias. Al descartar la injusticia y la arbitrariedad, podemos encontrar principios de la vida política que cada ciudadano puede justificar a los demás. Rawls escribe: «El objetivo es eliminar aquellos principios que sería racional proponer […] si supiéramos ciertas cosas que son improcedentes desde el punto de vista de la justicia» [7]. Sin embargo, cuando reconsidera su teoría, parece, como acabamos de ver, que abandona esta ambición.

> La cuestión es si la persona egoísta que imaginamos tras el velo de ignorancia está dispuesta a aceptar un principio de justícia porque es razonable o porque es racional

Podemos argumentar que el coste de alejarse de la elección racional y centrarse en la razonabilidad en su obra póstuma es precisamente aceptar que no *podemos* encontrar adecuadamente limitaciones formales generales desde las que obtener la sustancia de los principios políticos. No obstante, la ventaja distintiva de abandonar esta ambición es eliminar cualquier ambigüedad en la formulación fundamental del alcance de la teoría y centrarse en justificar a los demás los términos de la cooperación social que no sería razonable rechazar.

Combinación de condiciones de conducta razonable en un solo concepto

Lo que gobierna en la posición original son «premisas débiles, aunque ampliamente aceptadas» [8]. Este modo de deliberación y justificación de principios de justicia se resume en las últimas páginas de *Teoría*:

> «La justificación es un razonamiento dirigido a los que están en desacuerdo con nosotros, o a nosotros mismos cuando estamos indecisos […]. Al estar destinada a armonizarse mediante la razón, la justificación avanza a partir de lo que todos los individuos tienen en común».
>
> Rawls, J. (1999) A Theory of Justice. Revised edition. Cambridge

> Mass.: Harvard University Press, 508.

Para volver a un punto explicado arriba: todos los ciudadanos razonables podrán, por ejemplo, estar de acuerdo en que la discriminación racial es injusta, pero no tendrán tan claro cómo sería una distribución justa de la riqueza y la autoridad. Esto puede generar desacuerdos violentos. Ante la incertidumbre y los desacuerdos, la posición original representa condiciones que no son arbitrarias y rechazan la injusticia, y que todos podemos apoyar como razonables en nuestra reflexión. Nos permite prever diferentes concepciones de justicia y sus respectivos conjuntos de principios, empezando por premisas compartidas. Al usarla, podemos tomar nuestras propias decisiones sobre justicia deliberando con otras personas.

Mediante el razonamiento a partir de la posición original, formulamos principios que todos podemos encontrar justificados. Las partes no son personas de una sociedad justa ni personas de *nuestra* sociedad. Tan solo son las criaturas artificiales creadas para nuestra deliberación cuando *disentimos* sobre lo que conlleva la justicia, pero estamos motivados a encontrar principios de justicia que podamos justificarnos unos a otros.

El objetivo de subrayar la ambigüedad presentada por consideraciones de la elección racional es que debemos tener cuidado y no sobrestimar la función de la posición original en el argumento global de Rawls. Rawls no concibe la posición original como un método de descubrimiento de nuevos principios. La finalidad de la posición original es justificar principios que *ya* tenemos a nuestra disposición y que en el libro se formulan antes de la explicación del argumento de la posición original. En la construcción de la posición original, Rawls únicamente nos destaca la estructura relevante de las razones que ya tenemos en cuenta cuando nos referimos a la justicia.

¿Es esta lectura demasiado deflacionaria? Diría que no: es magnífico el atractivo de concebir la posición original como tan solo una heurística brillante para deliberar principios de una sociedad justa *empezando por* el valor que las personas dan al hecho de que sus instituciones sean justificables a los demás. Sin embargo, sí que plantea una pregunta más: ¿Por qué no presentar estos argumentos sin molestarnos en este experimento de pensamiento que plantea ambigüedades indeseables? Esta es precisamente la lección que T. M. Scanlon extrae de la observación citada arriba.

No obstante, no tenemos la obligación de seguir a Scanlon aquí y de abandonar la heurística. El argumento de Rawls recuerda que, al pensar sobre la justicia de nuestras instituciones políticas, nos enfrentamos a una dificultad, concretamente, para encontrar un punto de vista «que quede al margen de los rasgos y circunstancias particulares de la estructura básica existente y no quede distorsionado por ellos» [9]. Los principios de justicia son principios para la estructura básica. Sin embargo, nuestra estructura básica es injusta. Incluso cuando tenemos la motivación de encontrar principios e instituciones que sean justificables para los demás en la esfera política, hay algunas injusticias que nos parecen naturales. En otras palabras: incluso según la lectura deflacionaria anterior, la posición original es reveladora desde el punto de vista filosófico. Tal y como expone Rawls

justo al final de *Teoría*, la posición original nos permite concebir nuestro mundo social y nuestras instituciones políticas desde la perspectiva adecuada:

> «Podemos recordarnos a nosotros mismos que la naturaleza hipotética de la posición original da pie a la pregunta: "¿por qué deberíamos interesarnos por ella, sea moral o de otro tipo?" Recuerde la respuesta: las condiciones representadas en la descripción de esta situación son condiciones que, en realidad, ya aceptamos. O, si no lo hiciéramos, se nos puede persuadir para ello mediante consideraciones filosóficas de la clase que se introduce ocasionalmente.
>
> Puede darse una explicación de apoyo para cada aspecto de la posición original. Por consiguiente, lo que estamos haciendo es combinar en un concepto la totalidad de condiciones para las que estamos preparados tras la debida reflexión para reconocerlas como razonables en nuestra conducta respecto a los demás (§4).
>
> Una vez entendido este concepto, podemos fijarnos en el mundo social en cualquier momento desde el punto de vista requerido».
>
> Rawls, J. (1999) A Theory of Justice. Revised edition. Cambridge Mass.: Harvard University Press, 514.

Ante la elección de diferentes interpretaciones de la posición original, puede ser tentador suponer que debe preferirse la que entiende la elección racional como una parte esencial de la deliberación, porque nos ofrece una lectura más ambiciosa de la *Teoría* al expresar el objetivo de la parte básica de la ética colectiva en el egoísmo racional. Aquí he intentado describir brevemente las razones para suponer que la otra lectura es, en realidad, la más interesante y ambiciosa. Dejando a un lado la antigua esperanza de basar la moral en la prudencia, nos invita a ver lo reales que son los principios en los que debemos ponernos de acuerdo, tan solo sobre la base de encontrar un interés común en un desacuerdo razonable.

REFERENCIAS

1 — Rawls, J. (1999) *A Theory of Justice*. Revised edition. Cambridge Mass.: Harvard University Press, 20.
2 — Rawls, J. (2001) *Justice as Fairness: A Restatement*, edited by Kelly, E. Cambridge, Mass: Harvard University Press, 42-43.

3 — Rawls, J. (1999) *A Theory of Justice*. Revised edition. Cambridge Mass.: Harvard University Press, 102 [emphasis added].
4 — Ibídem, 5.
5 — Buchak, Lara, 'Taking risks behind the Veil of Ignorance, *Ethics*, April 2017.
6 — Scanlon, T. M. (2003) 'Utilitarianism and Contractualism', en *The Difficulty of Tolerance.* Cambridge: Cambridge University Press, 146 [cursiva añadida].
7 — Rawls, J. (1999) *A Theory of Justice*. Revised edition. Cambridge Mass.: Harvard University Press, 16-17.
8 — Ibídem, 18.
9 — Ibídem, 15.



Véronique Munoz-Dardé

Véronique Munoz-Dardé es profesora en el Departamento de Filosofía del University College of London y profesora adjunta Mills de Filosofía en la Universidad de California, Berkeley. Es conocida por sus trabajos sobre ética y filosofía política. En los últimos años, ha escrito artículos sobre la importancia de los números en el razonamiento práctico, sobre el ideal político de igualdad, sobre responsabilidad y sobre justicia distributiva. Es la autora de los libros *La justice sociale* (2001) y *Rawls: justice et critique* (2014).