PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE LETRAS Y CIENCIAS HUMANAS

El problema de la subdeterminación para el internalismo idealizado

Tesis para obtener el título profesional de Licenciado en Filosofía que presenta:

Sebastian Dasso Gallardo

Asesore(s):

Pablo Hernando José Quintanilla Perez-Wicht

Lima, 2024

Informe de similitud

Yo, Pablo Quintanilla Pérez Wicht, docente de la Facultad de Letras y Ciencias Humanas de la Pontificia Universidad Católica del Perú, asesor(a) de la tesis/el trabajo de investigación titulado: “El problema de la subdeterminación para el internalismo idealizado”, del/de la autor(a)/ de los(as) autores(as) Sebastian Dasso Gallardo, dejo constancia de lo siguiente:

-El mencionado documento tiene un índice de puntuación de similitud de 15 %.
Así lo consigna el reporte de similitud emitido por el software Turnitin el 03/09/2024 .

- He revisado con detalle dicho reporte y la Tesis o Trabajo de Suficiencia Profesional, y no se advierte indicios de plagio.

-Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lugar y fecha: Lima, 3 de septiembre de 2024

Apellidos y nombres del asesor: Quintanilla Pérez-Wicht, Pablo Hernando José	
DNI: 10276870	Firma
ORCID: 0000-0003-4588-3188	

Resumen

El presente trabajo tiene como objetivo presentar una nueva objeción contra el internalismo idealizado. Según el internalismo idealizado, tenemos razones para actuar conforme a lo que desearíamos en condiciones ideales (o a lo que nuestra contraparte ideal desearía que hagamos). Este trabajo sostiene que existen múltiples formas coherentes pero mutuamente excluyentes de idealizar los deseos de un agente, por lo cual el conjunto de deseos de un agente se encuentra racionalmente subdeterminado. En primer lugar, se argumenta que el internalismo idealizado no es capaz de encontrar un criterio no arbitrario para resolver el problema de la subdeterminación. Por un lado, los agentes actuales no pueden resolver el conflicto entre sus contrapartes ideales ya que no tienen acceso epistémico a ellas desde su perspectiva actual. Por otro lado, no es viable resolver los conflictos de contrapartes ideales utilizando promedios, ya que estos arrojan resultados arbitrarios cuando las preferencias de estas no están ordenadas de manera transitiva. En segundo lugar, se sostiene que el problema de la subdeterminación es un desafío significativo para el internalismo idealizado. Por un lado, desafía la fiabilidad del juicio de nuestras contrapartes ideales al ser un caso de desacuerdo entre pares epistémicos. Por otro lado, existe evidencia empírica que apoya la idea de que los deseos de, por lo menos, la mayoría de las personas se encuentran subdeterminados.

Palabras claves: racionalidad, internalismo de razones, racionalidad humeana, idealización

Abstract

This thesis aims to present a new objection against idealized internalism. According to idealized internalism, we have reasons to act according to what we would desire under ideal conditions (or what our ideal counterpart would desire us to do). This thesis argues that there are multiple coherent but mutually exclusive ways of idealizing an agent's desires, thus an agent's set of desires is rationally underdetermined. First, it argues that idealized internalism cannot find a non-arbitrary criterion to resolve conflicts between ideal counterparts. On the one hand, actual agents cannot resolve the conflict between their ideal counterparts since they do not have epistemic access to them from their current perspective. On the other hand, it is not feasible to resolve conflicts of ideal counterparts by using averages, as these yield arbitrary results when the preferences of these counterparts are not transitively ordered. Secondly, it argues that the problem of underdetermination poses a significant challenge to idealized internalism. On the one hand, it challenges the reliability of the judgment of our ideal counterparts as it is an instance of peer disagreement. On the other hand, there is empirical evidence supporting the idea that the desires of, at least, the majority of people are underdetermined.

ÍNDICE

1. Introducción

César está listo para cruzar el Rubicón. Marco Antonio le advierte que es una mala idea: la guerra civil será larga y sangrienta. Poco impresionado, César responde que esa consideración no le parece relevante. Indignado, Marco Antonio replica que incluso si no le interesa, salvar las vidas de soldados romanos es una consideración relevante. César pregunta qué razón podría tener para tomar en cuenta una consideración que no es de su interés: el hecho de que Marco Antonio valore las vidas de los soldados romanos es una buena razón en contra de pelear la guerra civil *para* Marco Antonio. ¿No es acaso poco razonable que una persona actúe según consideraciones que no son relevantes para ella? Frustrado, Marco Antonio termina la discusión recalcando que las razones de una persona no son puramente subjetivas: la vida de los soldados romanos es algo que todo general romano *debería* de estar interesado en preservar.

El desacuerdo entre César y Marco Antonio refleja una de las discusiones canónicas en las teorías de la racionalidad práctica [1]: ¿tienen las personas razones para hacer cosas que no quieren hacer? La bibliografía contemporánea suele frasear está pregunta en términos de estándares internos y externos: ¿son las razones de una persona internas o externas a sus deseos e intereses? (Wong 2006, Finlay y Schroeder 2017, Brunero 2017). Por un lado, la postura de César se asemeja al internalismo de razones, según el cual las razones de un agente son internas a su conjunto de deseos (Lord y Plunkett 2017). De ese modo, si un acto no se conecta de ninguna manera con los deseos de un agente, este no tiene ninguna razón para llevarlo a cabo. Por otro lado, la postura de Marco Antonio se asemeja al externalismo de razones, el cual sostiene que los agentes pueden tener razones para realizar actos independientemente de su

[1]Del mismo modo en que las teorías de la racionalidad epistémica tienen el objetivo de explicar lo que las personas tienen buenas razones para creer, las teorías de la racionalidad práctica tienen el objetivo de explicar lo que las personas tienen buenas razones para hacer (Resiner 2018, Wallace 2020, Kauppinen 2023). El uso del término "razón" suele variar de un autor a otro, por lo que su significado puede resultar oscuro (Fogal y Worsnip 2021). En el primer capítulo se aclaran los distintos usos que las principales teorías de la racionalidad práctica hacen de este término.

conjunto de deseos (Parfit 1997). Así, las razones de un agente pueden fundamentarse en un estándar externo, tal como el valor de las vidas de los soldados romanos.

En la bibliografía contemporánea, el internalismo idealizado es probablemente la postura más popular dentro de las teorías de la racionalidad práctica (Schroeder 2007, Sripada 2015, Sampson 2022). El internalismo idealizado sostiene que las razones de un agente se fundamentan en lo que este desearía en condiciones ideales (Sobel 2009). Por ejemplo, quizás César actualmente no desea salvar las vidas de soldados romanos porque se encuentra de mal humor. Sin embargo, en condiciones ideales de reflexión, César se daría cuenta de que en el fondo siente aprecio por sus soldados y sentiría una enorme culpa si su decisión de comenzar una guerra civil les costase la vida. Por lo tanto, César cuenta con razones para no cruzar el Rubicón. Uno de los principales atractivos del internalismo idealizado es que parece mantener un balance entre respetar los deseos de los agentes y corregir casos en donde estos fallan al reflexionar sobre sus deseos (Dorsey 2017).

El propósito de esta tesis es presentar una nueva objeción contra el internalismo idealizado. El núcleo de la objeción consiste en señalar que los deseos de los agentes se encuentran subdeterminados: existen múltiples formas mutuamente excluyentes de idealizar los deseos de, por lo menos, la mayor parte de personas. Si es correcto que los deseos de los agentes se encuentran subdeterminados, entonces los defensores del internalismo idealizado no cuentan con un método para decidir cuál de las múltiples contrapartes ideales es la correcta. Posteriormente, se sostendrá que dicha subdeterminación socava muchas de las motivaciones centrales detrás del método de la idealización.

El primer capítulo introduce el estado de la cuestión de las teorías de la racionalidad práctica. En particular, se caracterizará el debate entre internalistas y externalistas de razones y se analizarán las razones por las que los internalistas suelen adoptar el método de la idealización. El segundo capítulo presenta el argumento de la subdeterminación contra el internalismo

idealizado y explica por qué este presenta un desafío significativo. El tercer capítulo presenta respuestas a tres posibles objeciones contra el argumento de la subdeterminación.

2. Internalismo de razones y el método de la idealización

El objetivo de este capítulo es introducir el estado de la cuestión sobre el internalismo de razones. La primera sección caracteriza la naturaleza de las razones en las teorías de la racionalidad práctica. La segunda sección presenta el estado actual del debate entre internalistas y externalistas de razones. La tercera sección analiza la motivación detrás del uso del método de la idealización por parte de la mayoría de internalistas de razones.

2.1. Razones explicativas y razones normativas

Una distinción central en el estudio de la racionalidad práctica es la de razones explicativas y razones normativas. La distinción se puede apreciar en el siguiente par de oraciones:

(1) La razón por la que César asistió al senado el 15 de marzo es porque quería reunirse con Bruto

(2) César tenía una razón para no asistir al senado el 15 de marzo.

En primer lugar, (1) *explica* por qué César tomó cierto curso de acción. ¿En qué consiste dicha explicación? Una primera aproximación podría ser que estas razones hacen inteligible la conducta de César al mostrarnos desde su punto de vista qué consideraciones cuentan a favor de actuar de cierta manera (Raz 2009). Es claro que la consideración que César conscientemente tomó en cuenta para tomar la decisión de ir al senado el 15 de marzo fue *el hecho* de que tenía una reunión con Bruto, no *su deseo* de asistir a una reunión con Bruto. En general, los agentes suelen justificar sus actos en términos de estados de cosas y no en base a sus estados mentales *acerca* de estados de cosas (Heuer 2004). Esto ha llevado a algunos autores a sostener que las razones explicativas de la conducta de un agente se basan en hechos

externos a sus estados mentales (Dancy 1995). Si esto fuese correcto, deberíamos reformular (1) de la siguiente manera:

(1.ii) La razón por la que César asistió al senado el 15 de marzo es porque tenía una reunión con Bruto

El problema con (1.ii) es que no es claro cómo podría explicar la conducta de César en el caso de creencias falsas. Consideremos el siguiente ejemplo:

(3) La razón por la que César no trajo a sus guardias al senado es porque Bruto era inofensivo

Debido a que es falso que Bruto era inofensivo, es difícil ver qué hecho expresado por (3) podría explicar la conducta de César. Una lectura más natural de la situación sería la siguiente:

(3.ii) La razón por la que César no trajo a sus guardias al senado es porque *creía* que Bruto era inofensivo

Si esto es correcto, la manera de explicar la conducta de César es indagar en el rol causal de sus estados mentales al momento de realizar una acción. De ese modo, las razones explicativas de César en (1) serían los hechos psicológicos que explican su decisión de asistir al senado el 15 de marzo. Un ejemplo prominente de este modelo es la teoría de las razones como duplas de creencias y deseos de Donald Davidson (1963):

> R es una razón primaria por la que un agente performó una acción A bajo la descripción d solo si R consiste en un deseo del agente hacia acciones con cierta propiedad y una creencia del agente A, bajo la descripción d, tiene esa propiedad (687)

De ese modo, (1) expresa la dupla del deseo de César de reunirse con Bruto y la creencia de que la forma de satisfacerlo es asistir al senado el 15 de marzo. La necesidad de ambas condiciones se puede establecer mediante el análisis contra fáctico. Por un lado, si César no hubiese querido reunirse con Bruto, no habría asistido al senado el 15 de marzo. Por otro, si

César no hubiese creído que la forma de reunirse con Bruto era asistir al senado el 15 de marzo, tampoco habría asistido.

Algunos autores han resistido esta caracterización, dando pie a una extensa bibliografía sobre el debate entre las interpretaciones psicologistas y anti-psicologistas de las razones explicativas (Alvarez 2010, Hieronymi 2011, Wiland 2018)[2] . Dicha bibliografía se encuentra más allá de los límites de esta investigación. La distinción relevante para los fines de este trabajo es que las razones explicativas *describen* la conducta de un agente con la intención de hacerla inteligible.

En segundo lugar, (2) *prescribe* que hay una consideración que César racionalmente *debería* de tomar en cuenta [3] . La manera estándar de entender el tipo de razón expresado por (2) es que representa "una consideración a favor de actuar" (Scanlon 2004: 231) de cierta forma. Las razones normativas suelen ser caracterizadas como propiedades relacionales que establecen una conexión entre un hecho, un agente y un curso de acción (Alvarez 2018). Por ejemplo, el hecho de que Bruto ha planeado asesinar a César es una razón para que este no asista al senado el 15 de marzo.

A diferencia de las razones explicativas, las razones normativas no pueden estar fundamentadas por creencias falsas. Esto se debe a que las creencias falsas no pueden *justificar* por qué algo es una consideración a favor de realizar un acto[4]. Por ejemplo, la creencia de César de que Bruto era inofensivo es irrelevante para determinar si la conspiración de Bruto es una razón para que César no asista al senado el 15 de marzo. Una forma de interpretar la prescripción

[2]Una tercera postura consiste en diferenciar a las *razones explicativas* de las *razones motivacionales* (Alvarez 2010, Hieronymi 2011). Mientras que las primeras están constituidas por hechos psicológicos, las segundas corresponden con hechos externos.

[3]Por razones de brevedad, el término "razón" en este trabajo se referirá a razones normativas a menos que se indique lo contrario.

[4]Es crucial distinguir las razones subjetivas de las razones objetivas. Las razones subjetivas dependen de la perspectiva del agente: César podría tener buenas razones subjetivas para ir al senado si no había evidencia disponible sobre el complot de Bruto. Sin embargo, las razones normativas son razones objetivas: son consideraciones a favor de que un agente ejecute un acto independientemente de si este tenía evidencia sobre esas consideraciones (Fogal y Worsnip 2021).

expresada en (2) es que César racionalmente debería de tomar a la conspiración en su contra, de conocer su existencia, como una consideración en contra de asistir al senado el 15 de marzo.

¿Qué quiere decir que César racionalmente *debe* de hacer algo? Una respuesta simple sería sostener que expresa que César sería *irracional* si al enterarse de la conspiración en su contra decidiese no tomarla como una consideración en contra de asistir al senado el 15 de marzo [5]. El problema con esta respuesta es que no resuelve la duda de fondo: ¿por qué sería irracional no tomar en cuenta cierta consideración al momento de actuar? La clave para esclarecer esta pregunta es explicar en qué consiste que un agente sea racional. En la siguiente sección se explicará cómo esta tarea da origen a la división entre internalistas y externalistas de razones.

2.2. Razones internas y razones externas

En la escena contemporánea, las teorías de la racionalidad práctica suelen dividirse en dos campos.

(I) Internalismo de razones: S tiene una razón para Φ si y sólo si Φ contribuye a la satisfacción de alguno de deseos

(II) Externalismo de razones: S puede tener una razón para Φ incluso si Φ no contribuye a la satisfacción de alguno de deseos.

En primer lugar, el internalismo de razones sostiene que las razones de un agente son *internas* a su conjunto de deseos: si Φ no contribuye a la satisfacción de los deseos de S, entonces S no está *racionalmente* requerido a hacer Φ. ¿En qué se basa la asociación entre deseos y razones?

[5] ¿Hay una relación entre las razones normativas y las obligaciones morales? Externalistas de razones, tales como Derek Parfit (2011) y Peter Singer (2014), han defendido el racionalismo moral. Esta teoría sostiene que si S tiene la obligación moral de Φ, entonces S tiene una razón para hacer Φ. Muchos internalistas de razones rechazan el racionalismo moral: sostienen que no hay conexión necesaria entre que un acto sea moralmente condenable y que sea irracional (Foot 1972, Williams 1989, Street 2009, Manne 2014). Sin embargo, internalistas de razones tales como Michael Smith (1994) y Julia Markovits (2014) suscriben el racionalismo moral. Por lo tanto, la disputa entre internalistas y externalistas de razones es independiente de la disputa entre racionalistas y anti- racionalistas morales. Por ende, se dejará de lado cualquier discusión sobre el racionalismo moral: no se tomará postura sobre si un acto racionalmente permisible es necesariamente moralmente permisible.

David Hume famosamente dictaminó que "la razón es y debe ser la esclava de las pasiones y nunca puede pretender otro oficio que el de servirlas y obedecerlas" (T II.3.3 415). El internalismo de razones suele basarse en una teoría de la racionalidad práctica de corte humeano: un agente es racional en tanto sus acciones son un *medio* efectivo para cumplir con los *fines* dictaminados por sus deseos (Schroeder 2014). De ese modo, la racionalidad es de carácter "instrumentalista o hipotético" (Railton 2009 265) en tanto su rol no es determinar los fines que un agente debería de seguir, sino garantizar que los medios que siga sean coherentes con sus fines.

¿Por qué creer en el internalismo de razones? Hay un gran número de argumentos a favor de esta postura. Por razones de brevedad, se esbozarán las tres familias de argumentos más prominentes:

A. Razones y deliberación: quizás los argumentos más famosos a favor del internalismo se basan en la conexión que tiene el concepto de razón con nuestras prácticas ordinarias de deliberación (Finlay 2009). Por un lado, Bernard Williams (1979) sostuvo que S tiene una razón R para φ solo si es posible que S haga φ *debido* a R. De lo contrario, las razones de S serían distintas a las consideraciones que S reconocería como relevantes si deliberase sobre su conducta. Williams sostiene que si una razón R es completamente externa a las consideraciones relevantes para S, S no sería capaz de llegar a aceptar R a través de la reflexión y por ende no sería capaz de *actuar en base* a R. Por otro lado, Kate Manne (2014) sostiene que razonar con un agente implica tomar un punto de vista interpersonal: las consideraciones que intercambiamos con un interlocutor deben ser relevantes desde el punto de vista de ambos. Las razones externas no cumplen con este requisito, ya que son completamente ajenas a las preocupaciones, intereses o valores de nuestro interlocutor. Por lo tanto, Manne sugiere que las razones externas son intentos de "dar ordenes (...) o manejar la conducta" (2014:91) de nuestro interlocutor en vez de intentar *razonar* junto con él.

B. Razones y normatividad: si las razones normativas son propiedades relacionales, es necesario explicar por qué un hecho constituye una razón para un agente en particular.

El internalismo tiene una explicación simple: las razones de los agentes responden a consideraciones que *ellos mismos* reconocen como importantes. En contraste, explicar la normatividad de las razones externas enfrenta dos desafíos. Por un lado, el externalismo enfrenta *el problema de la autoridad*: si una razón está completamente desligada de los deseos de un agente, no es claro en virtud de qué esta debería de ser una consideración relevante para él (Finlay 2007). Shamik Dasgupta (2017) compara al externalismo con las teorías del mandato divino: ambas nos dicen que una autoridad demanda algo de nosotros , pero no son capaces de explicar *por qué* deberíamos de obedecer esas demandas. Por otro lado, el externalismo genera *el problema de la alienación normativa* al separar las razones de los agentes de sus propios valores y preocupaciones (Railton 1986, Rosati 1996). Esto se debe a que los agentes no pueden sentirse identificados con las demandas que las razones externas ejercen sobre ellos. Jack Samuel sostiene que este problema es producto de la forma en la que el externalismo aísla la normatividad de nuestras prácticas ordinarias de deliberación, lo cual transforma al concepto de razón en" algo extraño y distante, en vez de como algo que constituye nuestra experiencia" (2023: 161-162).

C. Parsimonia: la simpleza suele ser considerada como una virtud teórica: una teoría que postula menos entidades que las teorías rivales sin pérdida de poder explicativo es *pro tanto* preferible (Quine 1966). Por un lado, Mark Schroeder (2007) sostiene que el internalismo es más parsimonioso que el externalismo debido a que fundamenta todas nuestras razones prácticas en el mismo tipo de entidad. Ciertamente *algunas* de nuestras razones son internas a nuestro conjunto de deseos: es poco plausible sostener que César tiene una razón para tomar vino independientemente de que disfrute del vino o que este contribuya de alguna manera a sus intereses (Sobel 2005). Si la posición internalista funciona para algunas razones, ¿por qué pensar que no funciona para todas las razones? Por otro lado, el internalismo no requiere postular entidades distintas a las ya aceptadas por nuestras mejores teorías científicas (Lord y Plunkett 2017). Esto se debe a que fundamenta todas las razones prácticas en deseos, los cuales son entidades psicológicas ampliamente usadas en las ciencias para explicar la conducta humana (Ghoniem y Hoffman 2016).

¿Qué hace a un acto *irracional* según el internalismo de razones? Se suele sostener que el internalismo de razones caracteriza a la normatividad práctica en términos procedimentales: un

agente es irracional si es que delibera de manera incorrecta sobre sus deseos y creencias (Parfit 1997, Hooker y Streumer 2009). Por ejemplo, César sería irracional si es que desea vivir y sabe que asistir al senado le costará la vida, pero no desea no asistir al senado. La falla de deliberación de César consiste en su incapacidad de inferir una consecuencia obvia de sus creencias y deseos. Múltiples publicaciones recientes catalogan errores como el de César como fallas de *racionalidad estructural*: un agente es estructuralmente irracional si sus estados mentales son incoherentes entre sí (Fogal 2020, Worsnip 2021, Fink 2023, Lee 2024). Alex Worsnip caracteriza la incoherencia de estados mentales de la siguiente manera:

> Un conjunto de estados mentales actitudinales es simultáneamente incoherente si y sólo si es (parcialmente) constitutivo de los estados mentales en cuestión que, para cualquier agente que tenga estas actitudes, el agente está dispuesto, cuando se cumplen condiciones de máxima transparencia, a abandonar al menos una de las actitudes. Esto es, los agentes humanos están dispuestos a (al menos normalmente) no ser capaces (o al menos encontrar muy difícil) de sostener psicológicamente tal combinación de actitudes bajo condiciones de máxima transparencia (2018: 188)

¿En qué consisten las condiciones de máxima transparencia? Worsnip sostiene que estas se cumplen cuando un agente explícitamente cree que tiene un estado mental y no se presentan *defeaters* [6]tales como el autoengaño, la fragmentación mental, o cualquier falla relevante de autoconocimiento. Las condiciones de transparencia máxima funcionan como un criterio para formular una prueba contrafáctica: si se diesen, ¿sería el agente psicológicamente capaz de mantener cierta combinación de creencias? Worsnip sostiene que este criterio unifica muchas de nuestras intuiciones, aparentemente dispares, sobre requisitos de coherencia. Consideremos los siguientes ejemplos:

(A) Bajo condiciones de transparencia máxima, César no sería capaz de mantener la intención de matar a Pompeyo, creer que la única forma de hacerlo es zarpar a Egipto

[6]Un *defeater* es evidencia en contra de la fiabilidad de una creencia (Moretti y Piazza: 2018). Por ejemplo, la enorme propensidad de Napoleón al autoengaño es un *defeater* contra su creencia de que logrará escapar de Santa Helena y recuperará la corona.

y no tener la intención de zarpar a Egipto. Esto captura el requisito de coherencia entre fines y medios.

(B) Bajo condiciones de transparencia máxima, César no podría mantener la creencia de que prefiere montar caballo sobre tocar la lira, prefiere tocar la lira sobre beber vino y prefiere beber vino sobre montar caballo. Esto captura el requisito de transitividad de preferencias.

(C) Bajo condiciones de transparencia máxima, César no tendría la capacidad de mantener la creencia de que la anécdota contada por Pompeyo anoche es verdadera y la creencia de que todas las anécdotas contadas por generales romanos son falsas. Esto captura el requisito de coherencia entre creencias de primer orden y creencias de orden superior.

Por lo tanto, se puede sostener que el internalismo de razones concibe a la normatividad práctica en términos de *racionalidad estructural*. En el caso de (3), afirmar que César cuenta con una razón para llevar a sus guardias al senado implica que César sería irracional si no aceptase que cuenta con una consideración a favor de hacerlo[7]. La irracionalidad de César consiste en que negar que cuenta con una razón para llevar a sus guardias al senado es incoherente con su deseo de preservar su vida.

En segundo lugar, el externalismo de razones sostiene que las razones de un agente pueden ser *externas* a su conjunto de deseos. La característica central de las posturas externalistas es el rechazo a la tesis humeana de que solo las pasiones pueden determinar los fines que debemos seguir. Crucialmente, sostienen que tenemos razones para seguir ciertos fines que son *intrínsecamente* racionalmente requeridos (Parfit 2011, Shaffer-Ladau 2005, Singer 2012) ¿En qué consisten dichas razones? Muchos externalistas creen que las razones son un concepto

[7]Asumiendo, por supuesto, que César conoce la razón en cuestión (i.e. la conspiración de Pompeyo). Al igual que el criterio de incoherencia de estados mentales de Worsnip, esta caracterización de las razones se basa en una prueba contrafáctica: en caso la conociera, ¿sería irracional negar que tiene la razón en cuestión?

primitivo, por lo cual no es posible definirlas de manera no circular (Skorupski 2010, Scanlon 2014, Stratton-Lake 2018). El siguiente pasaje de Derek Parfit es representativo de esta postura:

> Es difícil explicar el concepto de una razón (...) Los hechos nos dan razones, podríamos decir, cuando cuentan a favor de que tengamos alguna actitud (...) Pero "cuenta a favor de" significa aproximadamente "da una razón para'". Al igual que otros conceptos fundamentales, como aquellos involucrados en nuestros pensamientos sobre el tiempo, la conciencia y la posibilidad, el concepto de razón es indefinible en el sentido de que no puede ser explicado solo con palabras. Debemos explicar tales conceptos de una manera diferente, haciendo que las personas consideren pensamientos que utilicen estos conceptos. Un ejemplo es el pensamiento de que siempre tenemos una razón para querer evitar estar en agonía (2011:31)

Tal como sugieren las últimas líneas del pasaje, los externalistas suelen adoptar la estrategia de apelar a nuestra intuición de que existen propósitos intrínsecamente irracionales, tales como desear vivir en agonía. ¿En qué consiste esta intuición? La intuición suele ser caracterizada como un "parecer intelectual" (Beeler 1998:211) según el cual una proposición nos parece verdadera exclusivamente en base a nuestro entendimiento de ella. Los externalistas defienden el valor evidencial de las intuiciones haciendo paralelos con su uso en las matemáticas (Clarke-Doane 2020), la lógica (Huemer 2005) y la epistemología (Cuneo 2007). Por ejemplo, se ha sostenido que nuestra intuición sobre la irracionalidad de desear vivir en agonía es tan fiable como nuestras intuiciones sobre los casos de Gettier (Russell 2017). Los externalistas suelen hacer extensivo uso de ejemplos hipotéticos sobre agentes que tienen propósitos intuitivamente irracionales[8]. Uno de los más célebres se encuentra en la obra temprana de Derek Parfit:

> "Cierto hedonista se preocupa mucho por la calidad de sus experiencias futuras. Con una excepción, le importa igualmente todas las partes de su futuro. La excepción es que tiene indiferencia de martes futuros. Durante todos los martes se preocupa de manera normal por lo que le está sucediendo. Pero nunca le importan los posibles dolores o placeres en un futuro martes. Por lo tanto, elegiría una operación dolorosa el martes siguiente en lugar de una

[8] Para una compilación y crítica al uso de dichos ejemplos, los cuales incluyen a un psicópata llamado Calígula, una mujer anoréxica y un hombre que dedica su vida a contar el césped, véase Street, S. (2009). "In defense of future Tuesday indifference: Ideally coherent eccentrics and the contingency of what matters". *Philosophical Issues* 19 (1), 273-298.

> operación mucho menos dolorosa el miércoles siguiente... Esta indiferencia es un hecho simple. Cuando está planeando su futuro, es simplemente verdad que siempre prefiere la perspectiva de sufrimiento extremo en un martes que el dolor más leve en cualquier otro día." (1987, 123-124)

¿En qué consiste la irracionalidad del hombre con indiferencia de martes futuros? Su desinterés por los martes futuros es coherente con sus otros estados mentales, por lo que no es claro en qué sentido su comportamiento sería *estructuralmente* irracional. La respuesta propuesta por los externalistas es que es *sustantivamente* irracional (Parfit 1997, Setiya 2007). Un agente es sustantivamente irracional si no responde de manera apropiada a la evidencia disponible (Worsnip 2018). Por ejemplo, sería sustantivamente irracional que César mantenga la convicción de que ganará la batalla de Gergovia mientras ve como su ejército es diezmado por los galos. Del mismo modo, el hombre con indiferencia de martes futuros es sustantivamente irracional al ignorar la evidencia proveniente de la intuición de que no podemos descontar nuestro sufrimiento en base a criterios arbitrarios, tales como la fecha que marca el calendario. ¿Es posible que la conducta de un agente sea simultáneamente *estructuralmente* racional y *sustantivamente* irracional? Supongamos, por ejemplo, que César cree que es el mejor general de toda la historia y que el mejor general de toda la historia no puede haber perdido una batalla[9]. Si al reflexionar sobre ambas creencias César concluye que jamás ha perdido una batalla, sus creencias serían coherentes entre sí y por ende estructuralmente racionales. Sin embargo, César ha perdido media docena de batallas. Creer que es un general invicto es irracional debido a que cuenta con evidencia abrumadora de lo contrario. Los externalistas proponen que la irracionalidad de César es similar a la del hombre con indiferencia de martes futuros: ambos cuentan con un conjunto coherente de creencias que va directamente en contra de la evidencia disponible. De ese modo, se rechaza la propuesta internalista de restringir la normatividad práctica a la coherencia entre fines y medios debido a que ignora la dimensión de la racionalidad que concierne la sensibilidad a la evidencia (Worsnip 2019).

[9]Este ejemplo es una adaptación de Worsnip y Fogal (2021)

Por lo tanto, se puede sostener que el externalismo de razones concibe a la normatividad práctica en términos de *racionalidad sustantiva.* Las razones son consideraciones que un agente debe tomar en cuenta en tanto responde de manera racional a la evidencia. De ese modo, la evidencia proveniente de nuestras intuiciones fundamenta las razones que los agentes tienen de forma independiente de su conjunto de deseos.

Adjudicar la disputa entre internalistas y externalistas de razones se encuentra más allá de los límites de esta investigación. El argumento central de este trabajo objeta el uso de la idealización por el internalismo de razones, pero esto no debe ser interpretado como una defensa del externalismo. Descartar el método de la idealización es compatible tanto con aceptar una versión no idealizada del internalismo como con abandonarlo en favor del externalismo. No se tomará postura sobre cuál de estas opciones es preferible.

2.3. Internalismo y el método de la idealización

En su versión más simple, el internalismo sostiene que S tiene una razón para Φ si Φ contribuye a la satisfacción de *cualquiera* de sus deseos actuales. El problema con el internalismo simple es que no es extensionalmente adecuado, pues produce resultados profundamente contraintuitivos al evaluar agentes con deseos defectuosos. Chris Heathwood (2005) clasifica a los deseos defectuosos en dos categorías: deseos producidos por mala información y deseos producidos por mal razonamiento.

En primer lugar, S puede desear Φ pero no tener la información necesaria para evaluar Φ. Bernard Williams (1979) ilustra esta idea con el caso de una mujer que toma de una botella de gin sin saber que esta contiene petróleo. Sin bien ella actualmente desea tomar el líquido de la botella, es claro que este deseo se extinguiría si supiese lo que realmente contiene. Por lo tanto, sería absurdo sostener que tiene buenas razones para tomar el líquido de la botella. Del mismo modo, el internalismo simple produce falsos negativos. Kate Manne (2016) sustenta este punto con la observación de que las personas suelen sentirse más cálidas cuando están cerca de morir

de hipotermia: muchas personas indigentes pierden la vida de esta manera. Claramente, sería absurdo sostener que cuando el deseo de buscar calor se extingue en una persona con hipotermia, esta ya no tiene razones para guarecerse.

En segundo lugar, el deseo de S por Φ puede ser estructuralmente irracional. Los ejemplos más discutidos sobre deseos estructuralmente irracionales son los escenarios que Sarah Buss llama "casos de auto alienación" (2013:18). En estos casos, los agentes tienen un deseo muy intenso que motiva su conducta, pero rechazan abiertamente este deseo y lamentan las conductas que produce. Quizás el ejemplo más célebre de este tipo es el adicto involuntario, presentando originalmente por Harry Frankfurt:

> [el hombre] odia su adicción y siempre lucha desesperadamente, aunque sin éxito, contra ella. Intenta todo lo que está a su disponibilidad para superar su deseo por las drogas. Pero estos deseos son demasiado para él e invariablemente lo conquistan. Él es un adicto involuntario, víctima de sus propios deseos (1971:19)

Los casos de autoengaño son otro ejemplo prominente de deseos estructuralmente irracionales (Scott-Kakures 1996). Existen dos familias de teorías sobre el autoengaño: el intencionalismo y el no intencionalismo. Según el intencionalismo, cuando un agente cree que p y cree que ~p, la creencia en ~p está *motivada* por la creencia en p (Davidson 1986). Este fenómeno suele ser explicado apelando a la partición psicológica de la mente: los agentes pueden aislar su creencia en p de modo que no tengan acceso consciente a ella mientras creen que ~p (Davidson 1982, Rorty 1988, Davidson 1982, Pears 1991). Según el no intencionalismo, los agentes que se auto engañan para creer ~p *no creen* que p, sino que *sospechan* que p y en respuesta desarrollan un sesgo sistemático que ignora o malinterpreta la evidencia disponible de modo que no se forme la creencia que p (Barnes 1997, Mele 2001).

Otros casos de deseos estructuralmente irracionales incluyen la debilidad de la voluntad (Davidson 1969, Mele 2010), las preferencias adaptativas (Elster 1981, Eftekhari 2021), las fobias (Korsgaard 1997, Sripada 2014) y las preferencias intransitivas, (Williamson 2024). En todos estos escenarios, el internalismo simple daría el veredicto intuitivamente incorrecto: los agentes que tienen deseos estructuralmente irracionales tienen buenas razones para satisfacerlos.

Mark Schroeder (2007) diagnostica el origen de estos contraejemplos en el hecho de que el internalismo simple es excesivamente permisivo: tendríamos razones para hacer casi cualquier acto si la valla es tan baja como promover la satisfacción de *cualquier deseo.* Esta dificultad es particularmente grave en vista de que los individuos suelen tener múltiples deseos cuya satisfacción puede resultar mutuamente incompatible. Schroeder llama a este desafío el *problema de las demasiadas razones*. Para ilustrar el problema, Schroeder pregunta si un agente tendría una razón para comer la batería de su auto. A todas luces la respuesta es no: este acto atenta contra su deseo de preservar su salud y no poner en riesgo su vida. Sin embargo, comer la batería contribuye a la satisfacción de al menos uno de sus deseos: aumentar su consumo de hierro. De acuerdo con el internalismo simple, el agente tendría una razón para ingerir la batería de auto en tanto logra satisfacer alguno de sus deseos: una conclusión intuitivamente absurda.

La estrategia que la gran mayoría de internalistas sigue para enfrentar el problema de las demasiadas razones es restringir el tipo de deseos que pueden fundamentar razones (Schroeder 2007, Sripada 2015, Sampson 2022)[10]. En particular, se sostiene que nuestras razones se fundamentan en los deseos que tendríamos *en condiciones ideales*. De ese modo, las razones

[10]Tanto Schroeder como Sampson son críticos del método de la idealización; sin embargo, reconocen que esta es la estrategia preferida por la mayor parte de los internalistas de razones.

de S no dependen de lo que S actualmente desea, sino de lo que S *desearía* si se encontrase en condiciones de reflexión ideales. Dado que S no tendría deseos defectuosos en condiciones ideales, el método de la idealización permite evitar los veredictos contraintuitivos del internalismo simple.

Los internalistas suelen justificar el método de la idealización con la observación de que parte de nuestras prácticas ordinarias de reflexión es aceptar que podemos estar equivocados sobre nuestros deseos (Dorsey 2017). En particular, se suele aceptar que hay circunstancias mejores que otras para reflexionar sobre nuestros deseos. David Sobel (2009) ilustra este punto haciendo una analogía con el sabor: nuestra apreciación del helado de frambuesa es subjetiva, pero aceptamos que hay circunstancias que pueden interferir con nuestra capacidad para disfrutarlo, como estar empalagados o tener pasta de dientes en la boca. En estas circunstancias desfavorables, nuestra impresión del helado de frambuesa no representa adecuadamente cómo lo apreciaríamos en condiciones normales. Del mismo modo, hay circunstancias que nublan nuestra reflexión: podemos estar cansados, intoxicados, abrumados por el estrés o conmovidos por un evento traumático. En casos como los anteriores, nuestro juicio no es representativo de la configuración de nuestro conjunto de deseos. Por lo tanto, el propósito de la idealización es situar a los agentes en las circunstancias más favorables posibles para poder reflexionar críticamente sobre sus deseos.

La versión más popular del método de la idealización es el modelo del consejero ideal, según el cual S tiene una razón para Φ si la contraparte ideal de S le aconsejaría Φ[11]. La contraparte

[11] ¿Por qué no se basan en lo que la contraparte ideal de S *desearía*? Supongamos que S es un fumador. Lacontraparte ideal de S jamás habría probado un cigarro. Por lo tanto, no tendría el deseo de *dejar* de fumar. En la bibliografía sobre el internalismo de razones, ejemplos como este se conocen como *el problema de la falacia condicional* (Johnson 1999). Virtualmente todos los autores internalistas discutidos hasta ahora aceptan el modelo del consejero ideal, entre ellos Williams (1979), Railton (1986), Smith (1994), Sobel (2009) y Manne (2014). Una excepción notable es Schroeder (2007).

ideal de S, S-i, comienza con los mismos deseos de S. El rol de S-i es reflexionar sobre los deseos que debería priorizar haciendo uso de toda la información relevante y las mejores capacidades cognitivas posibles. Después de la reflexión, el curso de acción que S-i le aconsejaría seguir a S determina las razones de S. El modelo del consejero ideal suele ser comparado con "el razonamiento del estilo *si yo fuese tú*" (Asarnow 2019:56). Es decir, cuando damos un consejo, es común tomar la perspectiva de nuestro interlocutor. Las contrapartes ideales están en condiciones idóneas para dar este tipo de consejos: tienen pleno acceso a nuestra perspectiva, pues toman como punto de partida nuestro conjunto de deseos.

¿En qué consisten las condiciones ideales? ¿Qué cláusulas las componen? Algunas de las propuestas más prominentes, en orden cronológico, son las siguientes:

Autor	Cláusula de información	Cláusula de racionalidad
Roderick Firth (1952)	Ser omnisciente	Ser desapasionado y racionalmente consistente
Richard Brandt (1954)	Saber toda la información empírica relevante	Representar la información de forma vívida y razonar de manera lógica
Bernard Williams (1979,1989)	Tener acceso a la información relevante	Poder comparar distintos escenarios mediante la reflexión imaginativa.
Peter Railton (1986)	Estar completamente informado sobre sí mismo y sus circunstancias	Representar la información de manera vívida. Estar libre de errores cognitivos y lapsos de racionalidad instrumental

David Lewis (1989)	Representar todos los escenarios relevantes con el mayor acceso imaginativo posible	
Dreier 1990		No sufrir de anormalidades afectivas y ser capaz de encontrar un balance entre las acciones que valora y las que se encuentra motivado a realizar
Smith (1994)	Tener información correcta sobre el mundo	Respetar los requisitos de coherencia de la racionalidad estructural
(Sobel 2009)	Tener información plena sobre el mundo y las opciones que son parte de su vida	Razonar correctamente
Manne (2013)	Ser un asesor bien informado y bien dispuesto	Tener la capacidad de razonar correctamente con el agente actual

Si bien cada propuesta cuenta con matices distintos, la mayoría de ellas converge en que las condiciones ideales involucran tener la información necesaria para evitar los deseos que son producto de la ignorancia y tener las capacidades de razonamiento necesarias para evitar los deseos estructuralmente irracionales. Un detalle crucial de estas propuestas es que suelen hacer

hincapié en la capacidad de explorar distintos escenarios con la imaginación con el fin de comparar cómo nos sentiríamos al satisfacer nuestros distintos deseos. La motivación de esta cláusula se explica claramente en el siguiente pasaje de David Lewis:

> Si alguien tiene poca noción de cómo sería vivir como un espíritu libre, sin ataduras de la ley, la costumbre, la lealtad o el amor (...) entonces, si valora ciegamente estas cosas o no, tiene poco que ver con si son sus verdaderos valores. Lo que le falta es un conocimiento imaginativo. Si tan solo pensara más profundamente e imaginara vívida y completamente cómo sería si estos supuestos valores se realizarán (y tal vez también cómo sería si no lo fueran), eso haría que su valoración fuera un indicador más confiable de un valor genuino. Y si pudiera obtener el conocimiento imaginativo más completo que sea humanamente posible, entonces, sugiero, su valoración sería un indicador infalible. Algo es un valor si y sólo si estamos dispuestos, en condiciones del conocimiento imaginativo más completo posible, a valorarlo." (1989:121) (énfasis añadido)

La observación clave de Lewis es que si estamos equivocados sobre cómo se sentiría una experiencia, es dudoso que genuinamente la deseemos. En el ejemplo de Lewis, si un agente desea ser un espíritu libre porque erróneamente lo imagina como una vida llena de alegría y libre de preocupaciones, es razonable pensar que no desea la *verdadera* experiencia de ser un espíritu libre, sino la experiencia que *él cree* es representativa de ser un espíritu libre. Por lo tanto, la capacidad de imaginar con un grado razonable de precisión cómo nos sentiríamos en distintos escenarios y comparar la satisfacción que sentiríamos en cada uno de ellos es fundamental para poder reflexionar sobre nuestros deseos.

La suposición central del método de la idealización es que nuestras contrapartes ideales tienen las herramientas necesarias para llegar a un veredicto sobre los deseos que deberíamos priorizar. Esta suposición parece acertada en los casos de deseos producidos por mala información o ciertos tipos de irracionalidad estructural. El siguiente capítulo argumentará que hay un tipo sumamente relevante de casos en los que falla.

3. El problema de la subdeterminación

El propósito de este capítulo es defender la tesis de la subdeterminación de los deseos y argumentar por qué esta presenta un desafío para el internalismo idealizado. La primera sección presenta el problema de la subdeterminación de los deseos. La segunda sección sostiene que este no puede ser resulto consultando a los agentes actuales debido a que no tienen acceso epistémico a sus contrapartes ideales. La tercera sección argumenta que la subdeterminación es evidencia de orden superior en contra de la fiabilidad del juicio de nuestras contrapartes ideales.

3.1. La subdeterminación de los deseos

Los internalistas de razones suelen resaltar la estrecha conexión entre deseo y motivación[12]. Según la caracterización estándar de los deseos, si S desea p y cree que Φ producirá p, entonces S tendrá la disposición de hacer Φ (Stalnaker 1984, Smith 1994, Sinhababu 2009). Por lo tanto, si S desea p, S estará dispuesto a reconocer que p es una consideración relevante debido a su inclinación a actuar conforme a p (Gibbard, 1990). Precisamente por el carácter distintivamente práctico de los deseos, los internalistas sostienen que estos pueden fundamentar el vínculo que las razones establecen entre un agente, un hecho y un curso de acción.

En la psicología cognitiva, el concepto de deseo suele ser asociado con el de recompensa (Paul y Cushman 2022). Una recompensa es un estímulo que refuerza una conducta, de modo que su presencia aumenta la probabilidad de realizar dicha conducta (White 2011). Los ejemplos clásicos de recompensas son la gratificación de necesidades fisiológicas, como alimentarse cuando se tiene hambre y dormir cuando uno está cansado. Estas recompensas son intrínsecas: cuando se producen, no las apreciamos como un medio para un fin, sino como un fin en sí mismas (Blain y Sharot, 2021)[13] Las recompensas intrínsecas no se restringen a procesos

[12]Dale Dorsey se refiere a esta conexión como "el corazón" del internalismo (2017: 196)

[13]Por supuesto, esto no implica que el sueño y la comida son recompensas en *todos los contextos* (Paul y Cushman 2022). Si un individuo no se encuentra cansado, no encontrará ninguna recompensa en dormir. Más aún, si un individuo ha dormido un largo número de horas, podría repelerle la posibilidad de seguir durmiendo.

fisiológicos: los agentes pueden encontrar gratificación intrínseca en la amistad, las creencias religiosas y el respeto de valores morales como la justicia y la compasión (Haidt y Joseph 2007).

Las recompensas intrínsecas contrastan con las recompensas instrumentales (Morris et al 2022). Las recompensas instrumentales son apreciadas en tanto permiten alcanzar algún tipo de recompensa intrínseca a largo plazo. Por ejemplo, el ejercicio tiene valor instrumental porque eventualmente puede resultar en la recompensa intrínseca de tener buena salud. Algunas recompensas instrumentales pueden ser altamente contextuales: puedo encontrar gratificante tomar un atajo que me permite llegar a tiempo al trabajo. Otras, como la acumulación de dinero, están arraigadas en nuestras prácticas sociales (Medvedev et al 2024). Lo que unifica a los distintos tipos de recompensas instrumentales es que dependen de la creencia de que su gratificación contribuirá a la satisfacción de un deseo intrínseco[14]. Por ejemplo, las personas eventualmente perderían la motivación de acumular dinero si este dejase de ser un medio para obtener bienes y servicios.

De esta discusión sobre los tipos de recompensa se pueden extraer dos clases de deseos:

- Deseos intrínsecos: S desea p intrínsecamente si su deseo por p no es un medio para satisfacer otro deseo.
- Deseos instrumentales: S desea p instrumentalmente si su deseo por p es un medio para satisfacer otro deseo.

Por lo tanto, debemos distinguir entre que un estímulo sea valorado intrínsecamente con que este sea valorado de forma independiente del contexto.

[14]Podría objetarse que las personas buscan el dinero sin tener otros deseos en mente. No es poco común que una persona tenga el deseo de ser millonaria sin tener una idea clara de lo quisiera hacer con ese dinero. Paul y Cushman (2022) explican este fenómeno apelando a la eficiencia cognitiva. Debido a que el dinero ha sido típicamente valioso instrumentalmente en el pasado, simplemente asumimos que sigue siendo valioso instrumentalmente en lugar de calcular de nuevo todas las formas en que uno podría alcanzar resultados intrínsecamente gratificantes al gastarlo.

El internalismo idealizado es efectivo para corregir deseos instrumentales. Esto se debe a que los deseos instrumentales dependen de *la creencia* de que serán un medio efectivo para satisfacer un deseo intrínseco. Por lo tanto, la idealización puede resolver conflictos entre deseos intrínsecos al corregir las creencias en las que estos dependen. Supongamos, por ejemplo, que Bruto desea apoyar las reformas de César y también desea apoyar a los optimates[15]. Si ambos son deseos instrumentales basados en el deseo intrínseco de mejorar la vida del pueblo romano, la contraparte ideal de Bruto puede acceder a la información necesaria para decidir cuál de las dos facciones es *el mejor medio* para conseguir el fin que Bruto intrínsecamente desea. De ese modo, el rol de la idealización es otorgar a los agentes la información y capacidades de reflexión necesarias para que consigan un conjunto coherente de deseos instrumentales.

Los casos centrales usados para ilustrar el método de la idealización pueden ser parafraseados en términos de deseos instrumentales:

(A) Vida nueva (Lewis 1989): un hombre tiene el deseo intrínseco de vivir una vida tranquila. Desea instrumentalmente vivir la vida de un artista bohemio debido a que cree que esta será la mejor forma de evitar el estrés laboral. Sin embargo, en condiciones ideales sabría que la incertidumbre laboral de un artista bohemio es enormemente estresante. Por lo tanto, el hombre no tiene buenas razones para convertirse en un artista bohemio.

(B) Vocación literaria (Railton 1986): Beth es una contadora que tiene el deseo intrínseco de tener una carrera en la que se sienta satisfecha. Beth tiene el deseo instrumental de dejar la contabilidad por la literatura debido a que cree que tiene una enorme capacidad para la escritura creativa. Sin embargo, en condiciones ideales Beth sabría que no tiene

[15]Los optimates eran la facción aristocrática del senado romano que se oponía a las políticas de César.

las aptitudes necesarias para escribir una novela exitosa. Por lo tanto, Beth no tiene buenas razones para dejar la contabilidad por la literatura.

(C) Relación fallida (Sobel 2009): Una mujer tiene el deseo intrínseco de encontrar el amor. Ella tiene el deseo instrumental de casarse con su novio de la secundaria debido a que cree que tendrán un matrimonio feliz. Sin embargo, en condiciones ideales sabría que su matrimonio está destinado a fracasar debido a que no son compatibles como pareja. Por lo tanto, la mujer no tiene buenas razones para casarse con su novio de la secundaria.

¿Es el método de la idealización efectivo para corregir deseos intrínsecos? Consideremos el siguiente ejemplo:

(D) Venganza en Dinamarca: Hamlet tiene el deseo intrínseco de vengar la muerte de su padre y el deseo intrínseco de proteger su vida. Hamlet sabe que su padre fue asesinado por su tío Claudio, quien ha usurpado el trono. Si Hamlet decide asesinar a Claudio, se expondrá a un gran peligro. Sin embargo, si decide no correr ese riesgo, tendrá que abandonar su resolución de vengar la muerte de su padre.

El método de la idealización ciertamente puede determinar cuál sería el curso de acción más efectivo para satisfacer cualquiera de los dos deseos intrínsecos de Hamlet. Este análisis se vería de la siguiente manera

(D-α): Hamlet tiene el deseo intrínseco de vengar la muerte de su padre. En condiciones ideales, sabría que Claudio se encuentra rodeado por guardias durante el día y jamás toca la comida antes de que sea probada por un séquito de catadores. De ese modo, la contraparte ideal de Hamlet, Hamlet (α), le aconsejaría infiltrarse en la alcoba de Claudio y apuñalarlo mientras duerme. Por lo tanto, Hamlet tiene una razón para apuñalar a Claudio mientras duerme.

(D-β): Hamlet tiene el deseo intrínseco de proteger su vida. En condiciones ideales, sabría que Claudio tiene planeado asesinarlo eventualmente, ya que considera muy riesgoso mantener vivo a un posible aspirante al trono. De ese modo, la contraparte ideal de Hamlet , Hamlet (β), le aconsejaría huir de Dinamarca esa misma noche. Por lo tanto, Hamlet tiene una razón para huir de Dinamarca cuando caiga la noche.

¿Cuál de las dos contrapartes ideales fundamenta las razones de Hamlet? Ciertamente no pueden ser ambas. De lo contrario, Hamlet contaría con razones contradictorias:

1. Hamlet tiene una razón para hacer lo que su contraparte en condiciones ideales le aconsejaría
2. Hamlet α es la contraparte ideal de Hamlet
3. Hamlet α le aconsejaría a Hamlet que apuñale a Claudio mientras duerme
4. Hamlet β es la contraparte ideal de Hamlet
5. Hamlet β le aconsejaría a Hamlet que no apuñale a Claudio mientras duerme y que escape de Dinamarca
7. Por lo tanto, Hamlet tiene una razón para apuñar a Claudio mientras duerme (1,2,3)
8. Por lo tanto, Hamlet tiene una razón para no apuñalar a Claudio mientras duerme (1,4,5)
9. Por lo tanto, Hamlet tiene una razón para apuñalar a Claudio mientras duerme y tiene una razón para no apuñalar a Claudio mientras duerme (7,8) [16]

El desafío que enfrenta el internalismo idealizado es que no es claro qué *criterio* se podría usar para determinar cuál de las dos contrapartes ideales de Hamlet es la correcta. Una objeción

[16] ¿Por qué una conjunción en vez de una disyunción? Como se puede ver en el primer capítulo, el método de la idealización no busca razones normativas *pro tanto*, sino que son razones normativas en vista de todas las consideraciones (*all things considered reasons*) (Sripada 2015).

común contra el internalismo consiste en sostener que no es posible justificar el uso de condiciones ideales sin apelar a criterios externos a los deseos de los agentes (Millgram 2000, Lillehammer 2000, Ripstein 2010). En particular, se sostiene que existen distintas cláusulas que podrían ser incluidas dentro de las condiciones ideales. De ese modo, no es claro de qué manera los internalistas podrían justificar la inclusión o exclusión de ciertas cláusulas sin un criterio externo que guíe lo que esas cláusulas *deben* capturar. Simon Ripstein ilustra claramente esta objeción haciendo una analogía con las teorías disposicionales de la percepción:

> Así como las afirmaciones sobre lo que un agente tiene razones para hacer deben ser defendidas apelando a lo que él o ella elegiría bajo circunstancias apropiadas, así también el empirismo clásico buscaba tratar las propiedades perceptuales y los objetos físicos como construidos a partir de cómo las cosas se verían para una persona bajo circunstancias perceptuales normales. Así como la fatiga, la desinformación o la agitación emocional pueden interferir con nuestras decisiones, también la mala iluminación puede interferir con nuestra percepción. (...) [Sin embargo] hablar sobre cómo se vería algo en condiciones normales requiere de una justificación de las condiciones normales. Pero "condiciones normales" debe especificarse en términos de objetos físicos. Especificar lo que es normal en términos de lo que se percibiría lleva a un regreso al infinito (...). La alternativa es reconocer que se requiere alguna explicación de la veracidad de las condiciones normales para que la percepción hipotética tenga cabida en una explicación del mundo físico (2020: 42-45)

La respuesta internalista suele consistir en señalar que las cláusulas que conforman las condiciones ideales se basan en criterios *internos* al agente: los mismos agentes reconocen que deberían corregir sus deseos cuando estos son el producto de información falsa o errores de razonamiento (Sobel 2009, Dorsey 2017, Ben-Moshe 2021). Sin embargo, esta respuesta al problema del criterio no es viable en el caso de conflictos de deseos intrínsecos. Debido a que los deseos intrínsecos son deseados *independientemente* de otras consideraciones, no es claro de qué manera el método de la idealización podría introducir un criterio interno que haga que los agentes reconozcan a uno de sus deseos intrínsecos como más importante que otro. [17]

[17] ¿No puede apelarse a la intensidad fenoménica con la que un agente siente un deseo? En el capítulo 1 se discutió el ejemplo del adicto involuntario, el cual repudia el deseo que siente con más intensidad. Por lo tanto, esta propuesto no es viable.

Eric Sampson (2022) ha recientemente formulado un argumento similar basado en la permisividad de la racionalidad estructural[18]. Debido a que la racionalidad estructural sólo exige que un conjunto de deseos sea coherente, si un agente tiene múltiples deseos, es posible construir múltiples contrapartes ideales que sean coherentes pero mutuamente excluyentes. Sampson sostiene que los internalistas no pueden justificar de manera no arbitraria la preferencia por una contraparte ideal sobre otra. Cuando los internalistas caracterizan el método de la idealización, suelen sostener que la reflexión permite resolver conflictos entre deseos. Sin embargo, esto no proporciona una explicación sobre *cómo* se resolvería dicho conflicto. Para ilustrar la forma arbitraria en la que procede la reflexión propuesta por el internalismo idealizado, Sampson propone imaginar como las distintas contrapartes ideales decidirían cuál de ellas debería fundamentar las razones de un agente:

> ¿Podría el comité haber decidido racionalmente algún otro consejo? Y la respuesta siempre parece ser "sí". Podrían haber hecho un sorteo para que un consejero ganara y pudiera dar el consejo[19]. (..) Podrían haber hecho una competencia de vencidas, o tirado dados, o jugado a piedra, papel o tijera (...). En resumen, podrían haber adoptado cualquier procedimiento imparcial para llegar a un acuerdo. Siempre y cuando decidieran emitir el consejo dado por al menos uno de los asesores, el consejo habría incorporado la perspectiva evaluativa del agente real tan bien como cualquiera de las alternativas. Pero esto significa que cualquier recomendación hecha por un comité es arbitraria (2022:107)

El núcleo del desafío que enfrentan los internalistas es que los deseos intrínsecos de los agentes están *subdeterminados*. En la filosofía de la ciencia, una teoría se encuentra subdeterminada cuando el mismo conjunto de observaciones que la confirma también podría confirmar una teoría alternativa (Stanford 2023).[20] Del mismo modo, los deseos de un agente se encuentran

[18]El argumento de esta tesis difiere con el de Sampson en dos puntos. Primero, el diagnóstico de Sampson no explica los casos en los que el método de la idealización funciona bien. Este trabajo lo explica apelando a la distinción entre deseos instrumentales e intrínsecos. Segundo, Sampson no desempaca la intuición de que elegir una contraparte ideal de manera arbitraria es indeseable. Este trabajo explica el desafío en términos de evidencia de orden superior (ver la tercera subsección del capítulo).

[19]Sampson usa "consejero ideal" para referirse a las contrapartes ideales.

[20]Quizás la formulación más famosa del problema es la de Bas van Fraassen (1980), según la cual existen teorías distintas que son empíricamente equivalentes, de modo que no hay evidencia posible que permita confirmar cuál de ellas es la correcta.

subdeterminados cuando pueden fundamentar distintas razones mutuamente excluyentes. Si los deseos de los agentes se encuentran subdeterminados, la idealización no es una respuesta efectiva al problema de las demasiadas razones. Esto se debe a que una de las motivaciones centrales del método de la idealización es *filtrar* los deseos de los agentes para evitar el resultado contraintuitivo de que los individuos tengan múltiples razones igualmente legítimas para realizar actos mutuamente excluyentes. Pero el beneficio teórico es elusivo: en vez de resolver el problema, solamente lo empuja del primer orden al segundo orden.

Podría objetarse que ejemplos como el de Hamlet son extremos y por no ende no representativos de los escenarios relevantes para el internalismo de razones. El caso de Hamlet podría ser caracterizado como un dilema moral genuino, en el cual se confrontan dos valores inconmensurables (Sinnot-Armstrong 1988). De ese modo, el problema de la subdeterminación sólo se produciría en un número muy restringido de casos.

En respuesta, los conflictos de deseos intrínsecos son extremadamente comunes en las decisiones de las personas ordinarias. Los tres ejemplos antes mencionados pueden ser reconstruidos para incluir conflictos entre deseos intrínsecos.

> **(A.ii)** Vida nueva: un hombre tiene el deseo intrínseco de vivir con poco estrés y el deseo intrínseco de ser autónomo. En condiciones ideales sabría que la vida de un artista bohemio es sumamente estresante, pero confiere un grado de autonomía mayor que el de cualquier otra ocupación. Por lo tanto, el hombre tiene una razón para ser un artista si desea intrínsecamente ser autónomo y no tiene una razón para convertirse en artista si desea intrínsecamente vivir con poco estrés.
>
> **(B.ii)** Vocación literaria: Beth es una mujer urarina que tiene el deseo intrínseco de obtener reconocimiento y el deseo intrínseco de honrar su cultura. En condiciones ideales, Beth sabría que la mejor forma de honrar su cultura es convertirse en escritora y publicar la primera novela en urarina, pero que su obra literaria jamás obtendrá

reconocimiento. Por lo tanto, Beth tiene una razón para convertirse en escritora si desea intrínsecamente honrar su cultura y tiene una razón para no convertirse en escritora si desea intrínsecamente obtener reconocimiento.

(C.ii) Relación fallida: una mujer tiene el deseo intrínseco de encontrar el amor y el deseo intrínseco de ser madre. En condiciones ideales, sabría que su único chance de tener hijos es casarse con su novio de la secundaria, pero que este matrimonio será conflictivo e inevitablemente terminará en el divorcio. Por lo tanto, la mujer tiene una razón para casarse con su novio de la secundaria si desea intrínsecamente convertirse en madre y no tiene una razón para casarse con su novio de la secundaria si desea intrínsecamente encontrar el amor.

Ninguno de los deseos en conflicto anteriores es atípico. El internalismo idealizado opera bajo la caracterización de que los agentes actuales son propensos a tener preferencias incoherentes: esta es la motivación central para fundamentar sus razones en los deseos que tendrían en condiciones ideales en vez de en los actuales. Por lo tanto, no debería sorprender a los internalistas que parte de las preferencias incoherentes de los agentes actuales sea tener constantes conflictos de deseos intrínsecos.

Incluso si los conflictos entre deseos intrínsecos fuesen infrecuentes, el método de la idealización aumentaría enormemente su probabilidad. Esto se debe a que las contrapartes ideales están expuestas a un número potencialmente ilimitado de opciones, tal como se ve en el siguiente pasaje de David Sobel:

> Supongamos que amaría el sabor de las piñas si las probase, pero actualmente no tengo el deseo de hacerlo. Mi actual falta de deseo por comer piñas no implica que no me beneficiaría comerla. La satisfacción de un deseo actual no parece correlacionarse con lo que es bueno para uno. Los deseos informados parecen ser mejores para esa tarea (2009: 336)

Es claro que nuestras contrapartes ideales estarían expuestas a información que va mucho más allá del sabor de la piña: sabrían cómo cada vida posible puede contribuir a la satisfacción de alguno de nuestros deseos intrínsecos. Si esto es el caso, cada deseo intrínseco puede generar un número enorme de razones que los agentes actuales jamás habrían considerado. Por ejemplo, Hamlet podría tener decenas de deseos intrínsecos distintos al de vengar la muerte de su padre y preservar su vida, cada uno de los cuales requeriría diferentes cursos de acción para ser satisfechos. La probabilidad de que al menos *algunos* de estos cursos de acción sean mutuamente excluyentes es extremadamente alta. Por lo tanto, el mismo método de la idealización tiende a forzar conflictos entre deseos intrínsecos.

3.2. El problema del acceso epistémico

Una posible respuesta al problema de la subdeterminación es apelar a los deseos de los agentes *actuales*. Según esta respuesta, los agentes actuales tienen la potestad de elegir cual de sus contrapartes ideales es preferible. Esta versión del internalismo de razones podría formularse de la siguiente manera:

- S tiene una razón para **Φ** si su contraparte ideal *preferida* desearía que haga **Φ**

El atractivo de (IV) se basa en que mantiene un balance entre mantener la estrecha conexión entre las preocupaciones de los agentes y sus razones y la capacidad de corregir la conducta irracional. Por un lado, en los casos en los que solo un curso de acción es estructuralmente racional, las razones de S se basan en lo que su *única* contraparte ideal desearía que haga. Por otro, en los casos en donde hay múltiples formas coherente pero mutuamente excluyentes de satisfacer los deseos de S, S tiene la potestad de determinar *desde su perspectiva actual* cuál de estas prefiere.

¿Cómo se aplicaría (III) en el caso de Hamlet? La contraparte ideal en la que se basan las razones de Hamlet es la que el Hamlet *actual* prefería tras reflexionar sobre cada una de ellas. Dicha reflexión consiste en examinar en qué consistiría vivir como cada una de sus contrapartes ideales y decidir cuál de ellas se ajusta mejor al tipo de persona que desearía ser. Según este modelo, el rol central de la idealización es determinar qué cursos de acción son estructuralmente racionales en vista de los deseos intrínsecos de un agente. Por ejemplo, determinaría que si Hamlet tiene el deseo intrínseco de salvar su vida, sería irracional que tome un bote en medio de una tormenta para huir de Dinamarca. En caso suceda un conflicto de deseos intrínsecos, y por ende existan múltiples contrapartes ideales mutuamente excluyentes, las razones de Hamlet se fundamentan en lo que este preferiría *desde su punto de vista actual.*

En los siguientes tres apartados se responderá a esta objeción. Primero, se sostiene que -por razones que los mismos internalistas aceptan- los juicios de los agentes actuales no son fiables. Segundo, se argumenta que incluso si lo fuesen, los agentes actuales no tienen acceso epistémico a las experiencias de sus contrapartes ideales. Tercero, se responde a dos objeciones en contra de las experiencias transformadoras que podrían ser capitalizadas por los internalistas.

3.2.1. La objeción de la irracionalidad

El primer problema con esta respuesta es que contradice el propósito inicial de la idealización. La razón por la que los internalistas suelen adoptar el método de la idealización es porque los agentes actuales cuentan con una pobre capacidad para reflexionar sobre sus deseos. ¿Por qué deberíamos de confiar en la evaluación que hacen sobre los méritos comparativos de sus contrapartes ideales?

Por ejemplo, este modelo sostiene que el Hamlet actual puede tomar una decisión razonable sobre cuál de sus contrapartes ideales es preferible. ¿Pero no es acaso el punto de postular

contrapartes ideales de Hamlet que el juicio del Hamlet actual no es fiable? La evaluación que el Hamlet actual haría sobre sus contrapartes ideales es susceptible a exactamente los mismos sesgos que el método de la idealización busca filtrar. Supongamos que el Hamlet actual decide que prefiere a la contraparte ideal que honra la promesa a su padre por sobre la que prioriza su seguridad personal. ¿No podría la elección de Hamlet estar influenciada por el autoengaño? Quizás el Hamlet actual recuerda el poco cariño que recibió de su padre, pero la creencia de que su padre jamás lo quiso le resultaría tan dolorosa que se encuentra reprimida[21] (Livingstone-Smith 2003). Si esto fuese el caso, Hamlet estaría eligiendo a la contraparte ideal que prefiere honrar la promesa hecha a su padre *sin tomar en consideración* la crucial pieza de información de que su padre jamás lo quiso. Por lo tanto, la elección del Hamlet actual sería irracional. Consecuentemente, este modelo llevaría a la conclusión absurda de que las razones de Hamlet están fundamentadas en un juicio irracional.

Quizás la observación más dañina para esta propuesta es que las decisiones influenciadas por sesgos cognitivos crean otro problema de subdeterminación. Walter Sinnot-Armstrong (2018) ha compilado numerosa evidencia empírica sobre la inestabilidad que tienen las intuiciones morales de las personas ordinarias. En particular, Sinnot-Armstrong sostiene que *el orden* en el que las personas consideran la información tiene un efecto significativo en sus juicios morales. Si bien el internalismo idealizado descarta la influencia del efecto marco en las decisiones de las contrapartes ideales, este todavía afecta los juicios que los agentes actuales hacen *acerca* de sus contrapartes ideales. Por ejemplo, la decisión del Hamlet actual cambiaría dependiendo de si considera primero a la contraparte que honra la promesa a su padre o a la que decide priorizar su seguridad. En ambos casos, la elección que el Hamlet actual haría es

[21] ¿De qué manera se encuentra reprimida? Según el intencionalismo sobre el autoengaño, Hamlet actual tiene la creencia de que su padre jamás lo quiso, pero no tiene acceso consciente a ella. De ese modo, "existen barreras entre partes de la mente "(Davison 1982: 228) que evitan que acceda a la creencia que ha reprimido. Según el no-intencionalismo, Hamlet actual cuenta con un sesgo que sistemáticamente ignora o malinterpreta la evidencia disponible de manera que no se forme la creencia de que su padre jamás lo quiso (Mele 2019).

profundamente contingente al momento específico en el que toma la decisión. De ese modo, el Hamlet actual tendría y no tendría razones para honrar la promesa hecha a su padre dependiendo del orden en el que considere la información. Por lo tanto, este modelo concluye otra vez que las razones de Hamlet se encuentran subdeterminadas.

3.2.2. La objeción de las experiencias transformadoras

El segundo problema con esta respuesta es que no es claro cómo los agentes actuales tendrían acceso epistémico a las perspectivas de sus contrapartes ideales. El desafío de imaginar la vida como una de nuestras contrapartes ideales surge de la enorme brecha entre sus experiencias y las nuestras. La razón de esto es que el problema de la subdeterminación surge de instancias en las que un agente está conflictuado entre varios deseos, cada uno de los cuales puede cambiar drásticamente el tipo de persona que es. Esto se ve claramente en los ejemplos usados por los internalistas para ilustrar el método de la idealización, los cuales incluyen luchar en una guerra (Williams 1979), dejar un trabajo seguro para seguir una carrera incierta como escritor (Railton 1986) y convertirse en un "espíritu libre sin ataduras por la ley, la costumbre, la lealtad o el amor" (Lewis 1989: 121). Experiencias como estas superan nuestras capacidades imaginativas debido a que nos demandan simular vivencias radicalmente nuevas y proyectar cómo sería vivir a largo plazo como una persona moldeada por ellas.

Por ejemplo, según este modelo sería razonable que Hamlet compare a sus dos contrapartes ideales considerando varios escenarios representativos de la vida de cada uno. Para poder imaginarlos, Hamlet debe ser capaz de reconstruir los elementos centrales que componen estos escenarios, muchos de los cuales serán de naturaleza experiencial (Paul 2015). ¿En qué consisten estos elementos experienciales? Algunos son respuestas fisiológicas viscerales, como el miedo paralizante que podría sentir al desenvainar su espada para enfrentar a Claudio. Otros son sentimientos nuevos, como el efecto compuesto de la soledad extrema después de vivir una década en el exilio. Finalmente, algunos podrían implicar cambios sensoriales dramáticos,

como la experiencia de pasar el resto de su vida en la oscuridad de una mazmorra en caso su plan sea descubierto. El desafío al que se enfrenta Hamlet es que comparar diferentes vidas requiere comprender la naturaleza subjetiva de experiencias desconocidas.

El segundo problema que enfrentaría Hamlet es que ambas de sus contrapartes podrían contar con preferencias que le resultan irreconocibles. La primera contraparte podría haber sido consumida por la venganza: su principal interés es descubrir a todos los co-conspiradores de Claudio y matarlos sin piedad. La segunda contraparte podría verse transformada por la soledad: décadas viviendo como un ermitaño lo ha hecho incapaz de soportar el contacto humano. Ambas contrapartes partieron de deseos que se encuentran en el Hamlet actual, pero la naturaleza de la decisión hizo que se conviertan en personas radicalmente distintas. ¿Cómo podría Hamlet comparar cómo se sentiría vivir con dos conjuntos de preferencias dramáticamente distintas? Crucialmente, no es claro si la decisión debería basarse en cuál de las dos contrapartes satisface mejor sus preferencias actuales o en base a cuál contraparte se sentiría más satisfecha en base a sus nuevas preferencias (Paul y Quiggin 2018). Por ejemplo, quizás la vida de un ermitaño no sea satisfactoria para el Hamlet actual, ¿cómo podría descartar que esa vida no le será muy satisfactoria desde la perspectiva de sus nuevas preferencias?

Laurie Paul (2014) llama a decisiones como las que enfrenta Hamlet *experiencias transformadoras*. El desafío de decidir si someterse a una experiencia transformadora es que son tanto epistémica como personalmente transformadoras Primero, tienen un carácter fenomenal tan distintivo que no podemos imaginar cómo sería vivir una experiencia transformadora antes de experimentarla. Segundo, producen un cambio dramático en nuestras preferencias, a menudo hasta el punto en que es difícil evaluar nuestras circunstancias futuras desde la perspectiva de nuestras nuevas preferencias. Paul menciona a las experiencias de convertirse en padre, recibir trasplantes cocleares y enlistarse en el ejército como ejemplos de

experiencias transformadoras. Otros autores han considerado a las conversiones religiosas (De Cruz 2018), los doctorados en humanidades (Rothman Abril 2013), y las sentencias de prisión (Lackey 2020) como experiencias transformadoras.

El objetivo original de Paul es desafiar a las teorías de elección racional basadas en utilidad subjetiva: ¿cómo podemos calcular la satisfacción que nos daría una experiencia si no podemos saber cómo nos hará sentir y no podemos saber cómo moldeará a nuestras preferencias futuras? Por razones muy similares, las experiencias transformadoras frustran el proceso de seleccionar racionalmente a una contraparte ideal. ¿Cómo podemos evaluar racionalmente sus experiencias si no sabemos cómo sería vivirlas y no sabemos cómo sería vivir bajo las nuevas preferencias que tendríamos como producto de estas vivencias?

Paul compara nuestra ignorancia ante las experiencias transformadoras con un muro epistémico: sabemos que lo que sucede más allá de él nos involucrará, pero "no sabemos cómo será ser ese yo" (2020:21). La autora sugiere que la insuperabilidad de este muro tiene sus raíces en la forma especial en la que adquirimos conceptos fenoménicos. Los conceptos fenoménicos son representaciones mentales cuyo rol es permitirnos reconocer, comparar y anticipar el carácter cualitativo de una experiencia (Chalmers 2007). Por ejemplo, el concepto fenoménico CAFÉ [22] me permite recordar el sabor del café que tomé anoche e imaginar el olor del café cuando lo veo en una vitrina. Los conceptos fenoménicos suelen ser distinguidos de los conceptos psicológicos (Chalmers 1996). Mientras que los primeros son conceptos de primera persona que nos permiten pensar acerca de cómo *se siente* una experiencia, los segundos son conceptos de tercera persona que codifican información usada en procesos cognitivos de orden superior (Machery 2009). De ese modo, el concepto psicológico CAFÉ me permite formular inferencias y hacer categorizaciones con respecto al café, pero no me da información sobre cómo *se siente* experimentar el sabor del café. Los conceptos psicológicos

[22]Se seguirá la notación de Laurence y Margolis (1999) que representa a los conceptos con letras mayúsculas.

pueden ser adquiridos de múltiples formas: una conversación o un manual corto pueden ser suficientes para que sea capaz de hacer clasificaciones competentes sobre qué bebidas cuentan como café.

¿Cómo se adquiere un concepto fenoménico? Existen múltiples teorías sobre la forma en la que los conceptos fenoménicos[23] se relacionan con su referente, pero la gran mayoría de ellas convergen en que es necesario vivir una experiencia para adquirir los conceptos fenoménicos relevantes (Balog 2009, Sundström 2011, Lee 2023[24]). El rol que cumple la experiencia es darnos acceso[25] directo a los elementos subjetivos de esa vivencia (Nida-Rumelin 1995, Chalmers 2003). De ese modo, la experiencia nos da percatación directa del olor del café, el sonido de un arpa y el sabor de Vegemite[26]. Para defender esta tesis, Paul cita el famoso *argumento del conocimiento* de Frank Jackson:

> Mary es una científica brillante que está, por alguna razón, forzada a investigar el mundo desde un cuarto blanco y negro (...) adquiere, supongamos, toda la información física disponible acerca de lo que sucede cuando vemos tomates maduros, o el cielo; y usa palabras como "rojo", "azul", etc. Ella descubre, por ejemplo, justo qué combinación de ondas electromagnéticas del cielo estimulan la retina; y exactamente cómo esto produce, a través del sistema nervioso, la contracción de las cuerdas vocales y la expulsión de aire de los pulmones que resulta en la pronunciación de la proposición "el cielo es azul". [...] ¿Qué sucederá cuando Mary sea liberada de su cuarto blanco y negro? (...) ¿Aprenderá algo nuevo o no? (Jackson 1982:132).

El uso que Paul hace del caso de Mary la neurocientífica podría confundir al lector. El propósito del argumento de Jackson es defender el dualismo de propiedades, según el cual las propiedades mentales son *ontológicamente* distintas a las propiedades físicas (Robinson 2020). Sin embargo, lo único que requiere la tesis de Paul es que el *contenido informativo* de los

[23]Por ejemplo, se ha sostenido que los conceptos fenoménicos son demostrativos internos (Horgan 1984), que están parcialmente constituidos por tokens de la experiencia a la que refieren (Papineau 2007) y que son conceptos de reconocimiento puros aplicados a la experiencia (Carruthers 2000, Tye 2003)

[24]Lee (2023) no cree que este requisito sea necesario, pero reconoce que su posición se encuentra en extrema minoría. En la siguiente subsección se discutirá la teoría de Lee.

[25]"Acceso directo" es mi traducción del término *acquaintance*, el cual proviene de la distinción hecha por Bertrand Russel (1911) entre conocimiento por descripción y conocimiento por *acquaintance*. *Acquaintance* suele ser caracterizado como la percatación de la verdad de una proposición de manera no inferencial (Farkas 2019).

[26]Vegemite es una pasta de untar australiana que cuenta con un característico sabor salado. Luego de ser usado como ejemplo por David Lewis (1988), se convirtió en un ejemplo típico en la bibliografía sobre conceptos fenoménicos (Gertler 1997, Alter y Walter 2007, Liu 2019).

conceptos fenoménicos sea distinto al de los conceptos no fenoménicos. La característica central de una experiencia transformadora es *epistémica*: no podemos acceder a la información necesaria para saber cómo nos hará sentir antes de vivirla. Esto es independiente de la pregunta ontológica sobre si la experiencia fenoménica está constituida por propiedades no físicas.

Los dualistas de propiedades y la mayor parte de fisicalistas[27] están de acuerdo en que nuestros conceptos fenoménicos están inferencialmente aislados de nuestros conceptos físicos: concuerdan en que los componentes cualitativos de la experiencia de ver rojo no pueden ser directamente inferidos de una descripción de las propiedades físicas del color rojo (Sundström 2011). Por un lado, los fisicalistas explican la brecha epistémica en base a diferencias en cómo los conceptos fenoménicos y no-fenoménicos se relacionan con su referente [28]. Por ejemplo, algunos autores sostienen que los conceptos fenoménicos funcionan como indexicales que apuntan directamente al contenido fenoménico de un estado mental (Perry 2001, O´Dea 2002, Prinz 2007). Por otro, los dualistas de propiedades sostienen que las características distintivas de los conceptos fenoménicos solo pueden ser explicadas por diferencias ontológicas entre las propiedades mentales y físicas (Chalmers 2006). Ambas explicaciones son compatibles con el muro epistémico de Paul: convergen en que no es posible adquirir los conceptos fenoménicos relevantes antes de vivir una experiencia.

El muro epistémico creado por las experiencias transformadoras hace que nuestra decisión de favorecer a una contraparte ideal sobre otras sea arbitraria. Si no tenemos acceso a los conceptos fenoménicos característicos de las experiencias de nuestras contrapartes ideales, carecemos de la evidencia necesaria para tomar una decisión informada. Esto socava el propósito de encontrar un criterio no arbitrario para solucionar el problema de la

[27]Por supuesto, no *todas* las posturas fisicalistas están de acuerdo con esto. Por ejemplo, el funcionalismo analítico rechaza que los conceptos mentales no puedan ser analizados en términos de conceptos no mentales (Lewis 1966, Phelan y Buckwalter 2012). Adjudicar los méritos de estas posturas está más allá de los límites de este trabajo.

[28]Esta postura no solo es compatible con el fisicalismo, sino que suele ser empleada como una estrategia defensiva contra el dualismo de propiedades. *La estrategia de conceptos fenoménico*s consiste en sostener que nuestras intuiciones dualistas son realmente el producto de diferencias en nuestros *conceptos* mentales y físicos (Loar 1990). Por ejemplo, se sostiene que la intuición kripkeana (1980) de que la conexión entre eventos físicos y mentales es contingente responde a las diferencias entre nuestros *conceptos* de ambos.

subdeterminación. Si el modelo nos lleva a tomar decisiones a ciegas, simplemente traslada el problema de la arbitrariedad a un nivel superior.

En esta sección se consideró la posibilidad de responder al problema de la subdeterminación apelando a los deseos de los agentes actuales. El desafió que las experiencias transformadoras presentan a esta propuesta tiene la siguiente estructura:

1. Si no podemos imaginar una experiencia, no podemos evaluarla racionalmente.
2. Si no poseemos los conceptos fenoménicos relevantes, no podemos imaginar una experiencia.
3. Si nuestras contrapartes ideales han experimentado experiencias transformadoras, no poseemos los conceptos fenoménicos relevantes sobre su experiencia.
4. Nuestras contrapartes ideales han experimentado experiencias transformadoras.
5. No poseemos los conceptos fenoménicos relevantes sobre su experiencia (3-4)
6. No podemos imaginar sus experiencias (2-5)
7. No podemos s evaluar sus experiencias racionalmente (1-6)

Debido a que el argumento es válido, la estrategia disponible para los internalistas es objetar alguna de sus cuatro premisas. Por un lado, la premisa 2 se basa en la definición estándar de los conceptos fenoménicos y la premisa 4 toma como punto de partida los ejemplos más prominentes para justificar el método de la idealización. Las premisas 1 y 3 son más contenciosas La siguiente subsección tendrá el objetivo de responder a objeciones contra ellas.

3.2.3. Dos objeciones a las experiencias transformadoras

La tesis de Paul sobre las experiencias transformadoras ha recibido múltiples críticas [29]. ¿Podrían los internalistas hacer uso de ellas para defender el método de la idealización? En esta sección se examinarán las dos objeciones más populares contra la teoría de Paul y se concluirá que no pueden ser usadas en favor del internalismo.

En primer lugar, se ha acusado a la teoría de Paul de conferir una importancia desproporcionada a la capacidad de imaginar los aspectos experienciales de una vivencia en la toma de decisiones (Barnes, 2015, McKinnon 2015, Moss 2018). Por ejemplo, Antti Kauppinen (2015) sostiene que no necesitamos imaginar cómo se siente una experiencia para determinar si esta nos traerá algo que valoramos intrínsecamente, tal como la amistad o la autosuperación. Sin embargo, esta respuesta no está disponible para los internalistas, ya que el método de la idealización hace uso extensivo de la capacidad de introspección imaginativa. Desde sus primeras versiones, los defensores del internalismo idealizado han enfatizado que imaginar vívidamente diferentes escenarios es crucial para deliberar de manera correcta (Williams 1979, Brandt 1979, Railton 1986). David Lewis incluso llegó a postular que "el conocimiento imaginativo es todo lo que necesitamos" (1989:123) para reflexionar adecuadamente sobre nuestros deseos.

El énfasis que el internalismo idealizado hace en la capacidad de introspección imaginativa responde a una dificultad que Connie Rosati llama *el problema de la apreciación* (1995:304). Dicho problema se origina en la distinción entre poseer información y haberla internalizado. Por ejemplo, consideremos a un hombre que nunca se ha tomado el tiempo para imaginar las consecuencias de ignorar la advertencia de su médico de cambiar su dieta después de ser

[29]Por ejemplo, *Transformative Experiences* (2014) dio lugar a un simposio en la revista *Philosophy and Phenomenological Research* (Paul 2015a) y una edición especial en *Res Philosophica* (Paul 2015b) en los cuales se escribieron más de una docena de ensayos críticos de la postura de Paul.

diagnosticado con diabetes. Supongamos que imaginar, incluso brevemente, la experiencia de enfrentar una amputación inducida por la diabetes convencería al hombre de cambiar su dieta. Claramente, *no imaginar* la experiencia sería reflexionar deficientemente. Esto se debe a que se estaría ignorando la información más motivacionalmente relevante: el profundo sufrimiento que le producirá ignorar la advertencia de su doctor. Por lo tanto, un elemento central de las condiciones ideales es asegurar que se internalice la información relevante haciendo uso de una imaginación vívida. Consecuentemente, bajo los mismos términos del internalismo idealizado, la reflexión que carece de conocimiento imaginativo es profundamente defectuosa.

En segundo lugar, se ha criticado a la teoría de Paul por ignorar que podemos extrapolar información de experiencias similares pasadas para imaginar cómo se sentiría una experiencia transformativa. Por ejemplo, puedo prever razonablemente cómo sería ser padre si anteriormente he cuidado a un pariente más joven (Harman 2015). En general, esta línea de respuesta enfatiza que las experiencias no deben entenderse como una unidad discreta, sino más bien como una colección de varios hechos fenomenales. Por ejemplo, la experiencia de ser padre comprende varios hechos fenomenales que se superponen sustancialmente con los de otras experiencias, tales como la experiencia de cuidar a un dependiente, experimentar estrés trabajando largas horas durante la noche y sentir orgullo por el éxito de un ser querido (Krishnamurthy 2015).

Esta objeción podría encontrar apoyo en la teoría gradual del conocimiento fenomenal de Andrew Lee (2023), la cual sugiere que nuestros conceptos fenoménicos varían en términos del grado de precisión con el que nos permiten familiarizarnos con una experiencia. El argumento de Lee apela a la estructura gradual de nuestras capacidades epistémicas para distinguir el carácter fenomenal de una experiencia. Por ejemplo, podemos entrenarnos para ser progresivamente mejores en distinguir el escarlata del carmesí. Crucial para la teoría de Lee es la noción de que cada experiencia consiste en un conjunto de propiedades fenomenales; los

conceptos fenomenales alcanzan un grado más alto de pureza a medida que se vuelven más efectivos en filtrar estas propiedades para que coincidan exactamente una experiencia. Por lo tanto, podemos poseer conocimiento aproximado de algo que no hemos experimentado directamente si previamente hemos encontrado experiencias que comparten una configuración similar de propiedades fenomenales. Por ejemplo, si tenemos experiencias del color carmesí, podemos obtener un acceso epistémico razonablemente preciso al color escarlata, ya que ambos conjuntos de propiedades fenomenales se intersecan significativamente.

En respuesta, mi argumento no es incompatible con las teorías graduales del conocimiento fenoménico. Si bien la teoría original de Paul requiere que las experiencias transformadoras pertenezcan a un nuevo tipo experiencial (Paul 2015a), lo único que el argumento de esta tesis necesita es que nuestro acceso epistémico a ellas sea extremadamente pobre. Según la teoría gradual de Lee tenemos acceso epistémico aproximado a cualquier experiencia fenoménica, incluyendo experiencias tan extrañas como "mover nuestro séptimo tentáculo y sentir un campo electromagnético" (Lee 2023:215). Lee defiende esta conclusión contraintuitiva argumentando que nuestro acceso epistémico a ellas es *extremadamente impuro*. Por ejemplo, nuestras experiencias tienen algo en común con las de los pulpos debido a que ambos somos seres que cuentan con un sistema nervioso; sin embargo, el parecido es tan distante que nuestros conceptos fenoménicos sobre las experiencias de los pulpos son *sumamente imprecisos.*

Por lo tanto, si las experiencias de nuestras contrapartes ideales son muy distintas a las nuestras, nuestros conceptos fenoménicos de ellas serán sumamente imprecisos. Si nuestros conceptos fenoménicos son muy imprecisos, la evidencia que tenemos sobre cómo se sentirían esas vivencias será muy pobre. Por lo tanto, este modelo estaría fundamentando nuestras razones en decisiones tomadas en base a evidencia sumamente pobre. Esto nos regresa a la objeción inicial: si el propósito del método de la idealización es corregir la baja fiabilidad de nuestros

juicios actuales, esta respuesta socava dicha motivación al fundar nuestras razones en una decisión hecha en base a evidencia poco confiable.

3.3. La subdeterminación como evidencia de orden superior

Una segunda respuesta podría consistir en negar que la subdeterminación sea un problema. Del mismo modo en el que en la ética normativa se diferencian los actos moralmente obligatorios de los moralmente permisibles (Harman 2015), podría sostenerse que existen cursos de acción racionalmente obligatorios y racionalmente permisibles. Así, las acciones mutuamente excluyentes que nuestras contrapartes ideales escogerían delimitan el rango de acciones racionalmente permisibles. Por lo tanto, los agentes tienen razones para hacer lo que *cualquiera* de sus contrapartes ideales prescribiría. Si esto fuese correcto, no habría necesidad de justificar nuestra preferencia por ninguna opción en particular: elegir cualquier contraparte ideal es igualmente razonable.

En respuesta, es un error ver a los consejos contradictorios como evidencia de que todas las opciones sugeridas son racionalmente permisibles. Todo lo contrario: si bien es cierto que cada curso de acción es favorecido por una contraparte ideal, también es cierto que son *desfavorecidos* por todas los demás. Lo que más debería preocupar a los internalistas es que en casos como estos solemos disminuir nuestra confianza en el juicio de todas las partes involucradas.

Consideremos, por ejemplo, a una persona daltónica que pregunta cuál es el color de su corbata a tres participantes de una conferencia. Si recibe tres respuestas muy distintas, sería razonable dudar de la fiabilidad de todas las respuestas y estar inseguro sobre el color de la corbata. Ciertamente no sería razonable asumir que las tres respuestas son razonables y conformarse con seleccionar la que más nos guste. Ejemplos como este suelen ser usados para argumentar que el desacuerdo entre pares epistémicos puede ser evidencia de orden superior en contra de la fiabilidad del juicio de las partes involucradas.

¿En qué consiste la evidencia de orden superior? Mientras que la evidencia de primer orden versa directamente sobre la credibilidad de una proposición, la evidencia de orden superior versa sobre la fiabilidad de las capacidades de formar juicios (Horowitz 2022). Por ejemplo, saber que tengo daltonismo es evidencia de orden superior en contra de la fiabilidad de mis juicios sobre el color de las cosas. Esto es, el hecho de que tenga daltonismo no implica que mi creencia de que estoy viendo una manzana roja sea falsa, pero debería de reducir significativamente mi confianza en ella.

¿En qué consiste el desacuerdo entre pares epistémicos? Se considera como pares epistémicos a dos agentes que cuentan con acceso a la misma evidencia y que cuentan con capacidades cognitivas similares (Croce 2021). En el ejemplo anterior, los tres participantes de la conferencia tienen acceso a la misma evidencia perceptual y no tenemos razones para dudar que tengan capacidades cognitivas similares. Por lo tanto, no tenemos un criterio para decidir cuál de los tres es la fuente más fiable.

La intuición detrás de este ejemplo es que, si dos pares epistémicos llegan a conclusiones diferentes a partir de la misma evidencia, al menos uno de ellos debe estar equivocado. Como ambos tienen la misma probabilidad de cometer errores, debemos disminuir nuestra confianza en la fiabilidad de ambos por igual. Yan Chen y Alex Worsnip formulan el argumento del desacuerdo entre pares epistémicos de la siguiente manera:

P.1 Si yo y mi par tenemos acceso a la misma evidencia y respondemos racionalmente, llegaremos a la misma actitud doxástica sobre p
P2. Mi par y yo no hemos llegado a la misma actitud doxástica sobre p
C1: Por lo tanto, o mi par y yo no contamos con la misma evidencia o al menos uno de los dos ha respondido irracionalmente a la evidencia (1-2)
P3: Mi par y yo compartimos la misma evidencia
P4: No hay razón para pensar que mi par es más propenso a responder irracionalmente que yo

C2: Por lo tanto, existe un chance de al menos 50% de que yo haya respondido irracionalmente a la evidencia sobre p. (en prensa: 8)

Por hipótesis, las contrapartes ideales son pares epistémicos: tienen las mismas capacidades de razonamiento y acceso a la misma información. Por lo tanto, la subdeterminación debería tomarse como evidencia de orden superior en contra de la fiabilidad de su juicio. En lugar de presentarnos un rango de acciones razonables, debería preocuparnos que cada una de nuestras contrapartes ideales tenga la misma probabilidad de estar *equivocada* sobre el curso de acción que deberíamos tomar.

Uno podría verse tentado a pensar que el argumento anterior se basa en alguna versión del conciliacionismo, una posición en la filosofía del desacuerdo según la cual *generalmente* debemos responder al desacuerdo entre pares epistémicos reduciendo la confianza en nuestras creencias (Ballantyne y Coffman 2012). Sin embargo, incluso aquellos que rechazan el[30] conciliacionismo aceptan que hay instancias en las que el desacuerdo entre pares es un *defeater* legítimo. La dialéctica entre conciliacionistas y sus críticos a menudo consiste en objetar que los argumentos del lado opuesto son una excepción dentro de un patrón más amplio y explicar de qué manera constituyen una excepción (Worsnip 2014). Si el desacuerdo entre contrapartes ideales cae dentro del subconjunto correcto de casos, el conciliacionismo y sus críticos convergerán en postular que es una instancia en la que el desacuerdo entre pares es un *defeater* legítimo.

Lo que hace excepcional al desacuerdo entre contrapartes ideales es que estamos comparando fuentes *externas* que estamos seguros son igual de confiables. Los conciliacionistas suelen favorecer ejemplos relacionados con fuentes externas a nosotros, como relojes (Christensen 2009) o termómetros (White 2009); incluso los críticos del conciliacionismo suelen estar de

[30]La postura opuesta al conciliacionismo se conoce como t*he steadfast view* (Ranalli y Lagewaard 2022). Según esta, *generalmente* debemos responder al desacuerdo entre pares manteniendo la confianza en nuestras creencias.

acuerdo con su veredicto en tales casos (Kelly 2010). La razón de esto es que la mayoría de los argumentos contra el conciliacionismo se basan en privilegiar *nuestra* perspectiva por sobre la de nuestro interlocutor. Por ejemplo, suelen invocar nociones de primera persona, tales como evidencia privada (van Inwagen 1996), confianza en nuestras intuiciones (Wedgwood 2010) o normas epistémicas centradas en el agente (Huemer 2011). Debido a que este es un caso de desacuerdo entre fuentes externas, no hay una manera directa para los internalistas de hacer uso de las objeciones al conciliacionismo para mantener la confianza en el consejo de nuestras contrapartes ideales.

Una opción disponible para los internalistas es negar que el problema de la subdeterminación sea análogo a los escenarios discutidos en la filosofía del desacuerdo. Esto podría justificarse afirmando que no hay una respuesta correcta sobre cómo reflexionar sobre nuestros deseos. A diferencia de casos como el de termómetros que muestran diferentes temperaturas, donde uno debe estar equivocado, la interpretación de deseos tiene un elemento subjetivo ineliminable. Dado que no hay criterios objetivos para evaluar y priorizar conjuntos de deseos, el desacuerdo frente a evidencia común no es evidencia de error.

En respuesta, esta postura es incompatible con el método de la idealización. Si no existe una forma correcta de reflexionar sobre los deseos de un agente, no es claro por qué y bajo qué criterios se deberían filtrar los deseos defectuosos. Si los internalistas están comprometidos a respetar nuestras prácticas ordinarias de deliberación, deben de tener en cuenta que normalmente la incertidumbre que experimentamos al enfrentar un dilema se debe a que creemos que hay *una opción correcta* y tememos elegir la incorrecta. Esta es la razón por la cual recibir consejos contradictorios a menudo tiene un efecto desestabilizador en lugar de reconfortante: no estamos contentos con tomar una decisión que tenga sentido para una persona razonable, queremos tomar *la decisión correcta.*

La importancia de establecer a la subdeterminación racional como un *defeater* es que ataca el propósito de la idealización. Cómo se ha visto en la anterior sección, el método de la idealización pretende corregir la poca fiabilidad que tienen los juicios de los agentes actuales. Por lo tanto, si la subdeterminación tiene el efecto de hacer que los juicios de nuestras contrapartes ideales sean poco fiables, esto reduce significativamente el atractivo del internalismo idealizado.

4. Respuestas a objeciones

El objetivo de este capítulo es responder a tres posibles objeciones. La primera sección examina la objeción de la conciliación, según la cual el problema de la subdeterminación puede ser resuelto promediando las preferencias de nuestras contrapartes ideales. La segunda sección discute la objeción del valor externo, según la cual es posible resolver el impasse entre contrapartes ideales haciendo uso de razones externas. La tercera sección responde a la objeción empírica, según la cual los conflictos entre deseos intrínsecos son extremadamente improbables.

4.1. La objeción de la conciliación

Los internalistas podrían responder al problema de la subdeterminación haciendo uso de promedios. Podría sostenerse que, si un agente cuenta con múltiples contrapartes ideales, el curso de acción correcto es el que en promedio satisfaga los deseos del mayor número de ellas. Debido a que el uso de promedios integra las preferencias de *todas* las contrapartes ideales, evita el desafío de encontrar un criterio no arbitrario para preferir a una contraparte ideal sobre otra. ¿Cómo se vería esta propuesta en el caso de Hamlet?:

Hamlet cuenta con tres contrapartes ideales. Primero, Hamlet (α) prioriza su deseo de venganza, por lo cual decide matar a Claudio. Segundo, Hamlet (β) prioriza su deseo de seguridad, por lo cual decide huir de Dinamarca. Finalmente, Hamlet (γ) prioriza su deseo de poder, por lo cual decide pactar con Claudio con el fin de obtener influencia en su régimen. Por

un lado, es claro que (α) y (β) son más semejantes entre sí que con (γ), la cual está dispuesta a comprometer sus valores en la búsqueda de poder. Por otro, debido a que (γ) se encuentra motivada por el egoísmo, es razonable pensar que priorizaría el deseo de seguridad sobre el deseo de vengar a su padre.

De eso modo, los deseos de cada una de las contrapartes se promediarían de la siguiente manera:

	Matar a Claudio	Huir de Dinamarca	Unirse a Claudio
Hamlet (α)	1	0	-1
Hamlet (β)	0	1	-1
Hamlet (γ)	-1	0	1
Promedio	0	0.3	-0.3

En respuesta, el problema con esta objeción es que es vulnerable a los contraejemplos clásicos contra la agregación de preferencias en la teoría de elección social, tal como la paradoja de Condorcet (List 2022). Según esta paradoja, preferencias individuales racionales pueden producir una agregación irracional cuando se configuran de forma intransitiva. (Van Deemen 2014). Supongamos que las preferencias de las contrapartes de Hamlet se hubiesen configurado de la siguiente forma:

	Matar a Claudio	Huir de Dinamarca	Unirse a Claudio
Hamlet (α)	1	0	-1
Hamlet (β)	0	-1	1
Hamlet (γ)	-1	1	0
Promedio	0	0	0

No hay nada individualmente irracional en las preferencias de cada contraparte; sin embargo, el resultado final arroja una agregación intransitiva. Por lo tanto, genera tres promedios idénticos: ningún curso de acción es colectivamente más favorecido que otro. Por lo tanto, no es claro qué criterio no arbitrario se podría usar para decidir cuál agregación es preferible sobre las otras. Consecuentemente, las agregaciones se encuentran *subdeterminadas,* con lo cual se estaría volviendo a caer en el problema que esta objeción buscaba resolver.

Incluso si la agregación final es transitiva, es posible que el resultado final deje a todas las contrapartes insatisfechas. Consideremos el siguiente ejemplo:[31]

	A	B	C	D	E	F	G	H
Contraparte 1	6	7	8	5	4	3	2	1
Contraparte 2	6	5	4	3	2	1	7	8
Contraparte 3	6	8	5	7	4	3	2	1
Contraparte 4	6	1	2	3	4	5	7	8
Contraparte 5	6	4	3	2	1	7	8	5
Contraparte 6	6	2	8	3	4	5	7	1
Contraparte 7	6	3	2	4	5	7	1	8
Contraparte 8	6	5	7	8	4	3	2	1
Promedio	6	4.37	4.8	4.37	3.5	4.25	4.5	4

Leyenda: Las preferencias de cada idealización se encuentra ordenas desde el 8 (la opción más deseada) hasta la 1 (la menos deseado)

[31]Debido a que este ejemplo usa 8 variables, se prefirió mostrar directamente los números. Ilustrarlo con un ejemplo hipotético no haría el caso más caso, sino que confundiría al lector.

En este caso, la opción A cuenta con el promedio más alto a pesar de que no es la opción preferida por ninguna de las contrapartes. De hecho, no es ni siquiera la segunda opción preferida por alguna de ellas. Por otro lado, la opción H es la preferida por 3/ 8 contrapartes. ¿Por qué asumir que las contrapartes estarían más satisfechas si se eligiese la opción A en vez de la opción H? ¿Por qué asumir que estarían más satisfechas con la opción con la media más alta en vez de con la opción con la moda más alta? En general, mientras más opciones se presenten, será menos claro cuál es el criterio adecuado para agregar las preferencias de las contrapartes ideales, lo cual vuelve a enfrentar al internalismo idealizado con el problema de encontrar un criterio no arbitrario para resolver la subdeterminación.

4.2. La objeción del valor externo

Una segunda objeción consiste en resolver el problema de la subdeterminación utilizando un valor externo que permita comparar a las contrapartes ideales. Por ejemplo, se podría apelar a las normas prudenciales, según las cuales tenemos razones para promover nuestro propio bienestar (Nagel 1970, Fletcher 2019, Sagdhal 2022). De este modo, cuando dos deseos intrínsecos se encuentren en conflicto, los agentes deben seguir el que mejor contribuya a su autopreservación. En el caso de Hamlet, se daría el veredicto de que, dado que Hamlet está conflictuado, debe priorizar su bienestar personal y, por ende, desistir de su deseo de vengar a su padre.

En respuesta, esta objeción derrota el propósito del internalismo al conceder que existen razones externas. Si las razones prudenciales son un valor externo, entonces deben tener autoridad sobre los agentes *incluso* si estos no tienen ningún interés por su propio bienestar (Worsnip 2018). Si esto es correcto, los internalistas tendrían que conceder que existen fines racionalmente requeridos, tales como la autopreservación. ¿Por qué es la autopreservación un fin racionalmente requerido? Los defensores del internalismo idealizado probablemente tendrían que apelar a nuestras intuiciones de que un agente racional no puede estar completamente desinteresado por su bienestar. Sin embargo, en ese caso se estaría concediendo

la estrategia argumentativa central del externalismo y rechazando los argumentos fundacionales del internalismo en contra de las razones externas. Por lo tanto, esta objeción es incompatible con el internalismo.

3. La objeción empírica

Una tercera objeción consiste en negar la relevancia práctica del problema de la subdeterminación. Se podría sostener que la gran mayoría de las personas tiene rasgos de personalidad consistentes, por lo que priorizan ciertos deseos sobre otros siguiendo un patrón estable. Así, a diferencia de Hamlet, la mayoría de las personas tiene una inclinación clara hacia valorar la venganza o la autopreservación, lo cual puede deducirse fácilmente a partir de su comportamiento previo. Por ejemplo, mientras una persona con un historial de decisiones egoístas priorizaría la autopreservación, una persona con un historial de autosacrificio priorizaría la venganza. En general, cuando las personas ordinarias tienen dos deseos intrínsecos contradictorios, suele ser el caso que uno de ellos es claramente más representativo de la orientación general de su personalidad.

Según esta objeción, el argumento de la subdeterminación exagera enormemente el grado en el que las personas ordinarias cuentan con deseos contradictorios. Lejos de ser un conflicto insoluble, priorizar un deseo intrínseco sobre otro es una decisión relativamente simple para la mayor parte de personas. Si se encuentran en las condiciones de reflexión adecuadas, las personas darán veredictos consistentes con la orientación general de su personalidad. Por lo tanto, en el mundo real, las decisiones realmente difíciles para los agentes son las que involucran conflictos de deseos instrumentales, los cuales son idóneos para el método de la idealización. Por ende, la subdeterminación de los deseos intrínsecos no es un desafío significativo para el internalismo idealizado.

En respuesta, esta objeción se basa en una predicción empírica muy específica: los internalistas tienen la carga de prueba de demostrar que los conflictos de deseos intrínsecos son superficiales para la mayor parte de personas. Crucialmente, el programa de investigación situacionista ha producido múltiples estudios que ponen en duda esta predicción empírica (Ross y Nisbet 1991, Doris 1999, Harman 2000). La tesis central del situacionismo es el rechazo de los rasgos de personalidad globales. Los rasgos de personalidad globales cuentan con dos características (Funder 1991):

(1) Consistencia: las características de personalidad se manifiestan en una amplia gama de contextos. Por ejemplo, una persona que se comporta de manera honesta en sus relaciones interpersonales tiene una alta probabilidad de también ser honesto en otros contextos.

(2) Estabilidad: las características de personalidad perduran a lo largo del tiempo. Por ejemplo, es altamente improbable que alguien que muestre gran honestidad en una prueba se comporte con deshonestidad en la prueba del día siguiente.

La crítica situacionista a los rasgos de personalidad globales consiste en mostrar que la conducta moral de las personas cuenta con una enorme variabilidad en base a pequeñas diferencias ambientales. La mayor parte de situacionistas utilizan a la compasión como ejemplo de un rasgo enormemente variable [32], algunos de los experimentos más citados son los siguientes:

(A) Sujetos que encuentran dinero en una cabina telefónica tienen un 88% de probabilidades de ofrecer ayuda a un transeúnte al que se le han caído documentos de un folder. Los sujetos en el grupo de control solo ayudaron al transeúnte 4% de las veces (Isen y Levin 1972)

[32]Por ejemplo, es el rasgo más discutido en la obra de John Doris (1998, 2000) y Gilbert Harman 2000, 2009).

(B) Sujetos que escuchan el grito de una mujer en la habitación de al lado acuden a ayudarla el 70% de las veces si se encuentran solos. Si los sujetos se encuentran acompañados, solo ayudan a la mujer el 7% de las veces (Latané y Darley 1970).

(C) Seminaristas que están en camino a dar una charla sobre la parábola del buen samaritano ayudan a un transeúnte en apuros 63% de las veces si tiene tiempo de sobra. Los seminaristas que están muy cortos de tiempo solo ayudan al transeúnte 10% de las veces (Darley y Batson 1973).

¿Qué explica la enorme variabilidad de la conducta compasiva? La explicación situacionista es que las personas ordinarias están "evaluativamente fragmentadas" (Doris 1998: 509), de modo que pueden priorizar un valor en cierto contexto y priorizar un valor contrario en otros contextos. Doris ilustra el grado dramático en el que el carácter puede estar fragmentado analizando la conducta de Joseph Mengele y Oskar Schindler:

> Era capaz de ser extremadamente bondadoso con los niños, ganarse su agrado, traerles azúcar, pensar en pequeños detalles de sus vidas cotidianas y hacer cosas que nosotros genuinamente admiraríamos… y luego, a su lado … el humo de los crematorios y aquellos niños, mañana o en media hora iban a ser enviados ahí. Ahí es donde radica la anomalía (Linton 1986:337. Citado en Doris 2000:58.)

> Algunos investigadores han concluido que las actitudes de los rescatadores (..) indican la existencia de una personalidad altruista (..) De hecho, los rescatadores muestran enormes inconsistencias. Oskar Schindler salvó a más de mil judíos en Polonia de ser deportados y asesinados, pero también era manipulador, un bebedor empedernido, un mujeriego y un mercader de la guerra que no se distinguió particularmente antes o después de la guerra. Incluso hay casos de antisemitas de toda la vida que se convirtieron en rescatadores (Doris 2000: 59)

Si el diagnóstico situacionista es correcto, la objeción empírica al problema de la subdeterminación no es viable. Esto no solamente se debe a que no podemos predecir la conducta heroica de Schindler en base a sus decisiones previas, sino a que es posible que *él mismo* se haya visto sorprendido por sus deseos nobles. Así como Schindler- o Hamlet-, las

personas ordinarias tienen conjuntos de deseos sumamente complejos: reflexionar sobre ellos no es una tarea trivial.

5. Conclusiones

En el presente trabajo se defendieron dos tesis principales. Por un lado, se argumentó que el internalismo idealizado no cuenta con un método no arbitrario para solucionar el problema de la subdeterminación de los deseos. Tanto en el segundo como en el tercer capítulo se exploraron y rechazaron posibles soluciones al problema de la subdeterminación. Por otro lado, se sostuvo que el problema de la subdeterminación presenta un desafío significativo para el internalismo idealizado debido a que socava muchas de sus motivaciones centrales.

En primer lugar, se argumentó que el problema de la subdeterminación nace de la posibilidad de conflictos de deseos intrínsecos. Si bien el método de la idealización puede ser efectivo para esclarecer conflictos entre deseos instrumentales, no es capaz de resolver conflictos entre deseos intrínsecos sin socavar muchas de sus motivaciones centrales. Por un lado, apelar a los deseos de los agentes actuales fundamentaría las razones de los agentes en decisiones pobremente informadas. Por otro lado, utilizar promedios cuando se producen conflictos entre contrapartes ideales termina por arrojar resultados arbitrarios. Finalmente, apelar a un valor externo concede las tesis centrales del externalismo de razones.

En segundo lugar, se sostuvo que el problema de la subdeterminación no puede ser ignorado debido a que presenta un desafío significativo para las tesis centrales del internalismo idealizado. Por un lado, los conflictos entre contrapartes ideales son una instancia de desacuerdo entre pares, lo cual cuenta como evidencia de orden superior en contra de la fiabilidad del juicio de todas las partes involucradas. Por otro lado, los conflictos de deseos

intrínsecos son sumamente comunes en las personas ordinarias, lo cual hace inviable decir que el problema de la subdeterminación se confina a casos marginales.

Crucialmente, los argumentos de esta tesis no deben ser necesariamente interpretados como una crítica al internalismo de razones en general (y, por lo tanto, como una defensa indirecta del externalismo de razones). Esto se debe a que el problema de la subdeterminación solo se produce cuando se emplea el método de la idealización. Por lo tanto, el argumento de esta tesis es compatible tanto con el externalismo de razones como con versiones no idealizadas del internalismo de razones. Este trabajo no toma postura sobre cuál de estas dos alternativas es preferible.

6. Bibliografía

Alvarez, M. (2010). *Kinds of Reasons: An Essay in the Philosophy of Action*. Oxford: Oxford University Press.

Alvarez, M. (2018). "Reasons for action, acting for reasons, and rationality". *Synthese* 195 ,3293-3310.

Asarnow, S. (2019). "Internal Reasons and the Boy Who Cried Wolf". *Ethics* 130 (1),32-58.

Alter, T. y S. Walter (2007). "Introduction to phenomenal concepts and phenomenal knowledge". En Alter, T. y S. Walter (eds), *Phenomenal concepts and phenomenal knowledge*. Oxford: Oxford University Press, 1-14.

Ballantyne, N. y E. J. Walter (2011). "Uniqueness, Evidence, and Rationality". *Philosophers' Imprint* 11 (8),1-13.

Balog, K. (2009). "Phenomenal Concepts". En McLaughlin, B. y otros (eds.), *The Oxford Handbook of Philosophy of Mind*. Oxford: Oxford University Press, 292-312.

Bastien, B. y T. Sharot (2021). "Intrinsic reward: potential cognitive and neural mechanisms, Current Opinion". *Behavioral Sciences* 39,113-118.

Barnes, A. (1997). S*eeing through Self-Deception*. Nueva York: Cambridge University Press.

Barnes E. (2015). "What You Can Expect When You Don't Want to be Expecting". P*hilosophy and Phenomenological Research* 91(3), 775–786.

Ben-Moshe, N. (2021). "A Defense of Modest Ideal Observer Theory: The Case of Adam Smith's Impartial Spectator". *Ethical Theory and Moral Practice* 24 (2),489-510.

Brandt, R. (1954). “The definition of an "ideal observer" theory in ethics”. *Philosophy and Phenomenological Research* 15 (3), 407-413.

Brandt, R. (1979). *A theory of the good and the right*. Oxford: Claredon Press.

Brunero, J. (2017). “Recent Work on Internal and External Reasons”. *American Philosophical Quarterly* 54 (2), 99-118.

Buss, S. (2013). “The Possibility of Action as the Impossibility of Certain Forms of Self-Alienation'”. En Shoemaker, D. (ed.), *Oxford Studies in Agency and Responsibility* 1, 12-46.

Carruthers, P. (2018). “Basic questions”. *Mind and Language* 33 (2), 130-147.

Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory.* Nueva York: Oxford University Press.

Chalmers, D. (2002). “The content and epistemology of phenomenal belief”. En Jokic, A. y Q. Smith (eds.), *Consciousness: New Philosophical Perspectives.* Nueva York: Oxford University Press, 220-272.

Chalmers, D. (2006). “Phenomenal concepts and the explanatory gap”. En Alter, T. y S. Walter (ed.), *Phenomenal Concepts and Phenomenal Knowledge*. Nueva York: Oxford University Press, 167-194.

Chen, Y. y A. Worsnip (en prensa). “Disagreement and Higher-Order Evidence”. En Baghramian, M. y otros (eds.), *Routledge Handbook of Disagreement.* Nueva York: Routledge.

Christensen, D. (2009). “Disagreement as Evidence: The Epistemology of Controversy”. *Philosophy Compass* 4 (5), 754-767.

Croce, M. (2023). “The Epistemology of Disagreement”. *Routledge Encyclopedia of Philosophy*. https://www.rep.routledge.com/articles/thematic/the-epistemology-of-disagreement/v-1, consultado el 10 de Julio del 2024.

Dancy, J. (1995).” Why There Is Really No Such Thing as the Theory of Motivation”. *Proceedings of the Aristotelian Society* 95 (1):1-18.

Darley, J. M. y B. Latane (1968). “Bystander intervention in emergencies: Diffusion of responsibility”. J*ournal of Personality and Social Psychology*, 8(4), 377-383.

Darley, J. M. y C.D. Batson (1973). "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior”. *Journal of Personality and Social Psychology*, 27(1), 100–108.

Dasgupta, S. (2017). “Normative Non-Naturalism and the Problem of Authority”. *Proceedings of the Aristotelian Society* 117 (3), 297-319.

Davidson, D. (1963). “Actions, Reasons, and Causes”. *Journal of Philosophy* 60 (23), 685-700.

Davidson, D. (1982) “Paradoxes of Irrationality”. En Wollheim, R y J. Hopkins (eds), *Philosophical Essays on Freud.* Cambridge: Cambridge University Press, 289-305.

Davidson, D. (1986), “Deception and Division”. En Elster, J (ed.), *Problems of Rationality*. Oxford: Clarendon Press, 79–92.

De Cruz, H. (2018). “Religious Conversion, Transformative Experience, and Disagreement”. *Philosophia Christi* 20 (1), 265-276.

Doris, J. (1998). “Persons, situations, and virtue ethics”. *Noûs* 32 (4), 504-530.

Doris, J. (2002). *Lack of Character: Personality and Moral Behavior.* Nueva York: Cambridge University Press.

Dorsey, D. (2017).” Idealization and the Heart of Subjectivism”. *Noûs* 51 (1), 196-217.

Dreier, J. (1990). “Internalism and speaker relativism”. *Ethics* 101 (1), 6-26.

Eftekhari, S. (2021). “The Irrationality of Adaptive Preferences: A Psychological and Semantic Account”. *Utilitas* 33 (1), 68-84.

Elster, J. (1983). *Sour grapes: Studies in the subversion of rationality*. Cambridge: Cambridge University Press.

Farkas, K. (2019). “Objectual Knowledge”. En Raleigh, T. y J. Knowles (eds.), *Acquaintance: New Essays.* Oxford: Oxford University Press, 260-276.

Fink, J. (2023). “The Essence of Structural Irrationality”. *Journal of Ethics and Social Philosophy* 26 (2), 377-419.

Finlay, S. (2006). “The Reasons that Matter”. *Australasian Journal of Philosophy* 84 (1), 1 – 20.

Finlay, S. (2009). “The Obscurity of Internal Reasons”. *Philosophers' Imprint* 9:1-22.

Finlay, S y M. Schroeder (2017). “Reasons for Action: Internal vs. External”. En Zalta, E. (ed), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/fall2017/entries/reasons-internal-external, consultado el 10 de julio del 2024.

Firth, R. (1952). “Ethical absolutism and the ideal observer”. *Philosophy and Phenomenological Research* 12 (3), 317-345.

Fletcher, G. (2021). *Dear Prudence: the nature and normativity of prudential discourse*. Oxford: Oxford University Press.

Fogal, D. (2020). “Rational Requirements and the Primacy of Pressure”. *Mind* 129 (516), 1033-1070.

Fogal, D. y A. Worsnip (2021) “Which Reasons? Which Rationality?”. *Ergo* 8 (11), 306-343.

Frankfurt, H. G. (1971). “Freedom of the Will and the Concept of a Person”. *The Journal of Philosophy*, 68 (1), 5–20.

Funder, D. C. (1991). “Global Traits: A Neo-Allportian Approach to Personality”. *Psychological Science* 2 (1), 31-39.

Fürst, M. (2023). “Closing the Conceptual Gap in Epistemic Injustice”. *Philosophical Quarterly* 74 (1), 1-22.

Gertler, B. (1999).” A defense of the knowledge argument”. *Philosophical Studies* 93 (3), 317-336.

Gibbard, A. (1990). *Wise choices, apt feelings: a theory of normative judgment.* Cambridge: Harvard University Press.

Ghoniem A. y W. Hofmann (2016). “Desire”. En Zeigler-Hill, V. y T. Shackelford (eds), *Encyclopedia of Personality and Individual Differences.* Springer Cham. https://doi.org/10.1007/978-3-319-28099-8_501-1 , consultado el 10 de julio del 2024.

Haidt, J. y C. Joseph (2007). “The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules”. En Carruthers, P. y otros (eds.), *The innate mind: Vol. 3. Foundations and the future*. Nueva York: Oxford University Press, 367-392.

Harman, E. (2015). “Transformative Experiences and Reliance on Moral Testimony”. *Res Philosophica* 92 (2), 323-339.

Harman, G. (1999). “Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error”. *Proceedings of the Aristotelian Society* 99, 315-331.

Heathwood, C. (2005). “The problem of defective desires”. *Australasian Journal of Philosophy* 83 (4), 487- 504.

Heuer, U. (2004). “Reasons for actions and desires”. *Philosophical Studies* 121 (1),43–63.

Hieronymi, P. (2011). “Reasons for Action”. *Proceedings of the Aristotelian Society* 111 (3), 407-427.

Horgan, T. E. (1984). “Jackson on physical information and qualia”. *Philosophical Quarterly* 34 (2), 147-52.

Horowitz, S. (2022). “Higher-Order Evidence”. En Zalta, E. (ed), The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/fall2022/entries/higher-order-evidence, consultado el 10 de julio del 2024.

Huemer, M. (2011). “Epistemological egoism and agent-centered norms”. En Dougherty, T. (ed.), *Evidentialism and its Discontent*s. Oxford: Oxford University Press, 17-33.

Hume, D. (1978). *Treatise of human nature*. Edición de Selby-Bigge, L.A. Oxford: Oxford University Press.

Isen, A. M. y P.F. Levin (1972). “Effect of feeling good on helping: Cookies and kindness”. *Journal of Personality and Social Psychology* 21(3), 384-388

Kauppinen, A. (2023). “The Epistemic vs. the Practical”. En Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics* 18, 137-162.

Korsgaard, C. (1997) “The Normativity of Instrumental Reason”. En Cullity, G. y N. Gaut Berys (eds.), *Ethics and practical reason*. Nueva York: Oxford University Press, 27-68.

Kelly, T. (2010). “Peer disagreement and higher order evidence”. En Goldman, A. y D. Whitcomb (eds.), *Social Epistemology: Essential Readings.* Oxford: Oxford University Press, 183-217.

Kripke, S. (1980). *Naming and Necessity: Lectures Given to the Princeton University Philosophy Colloquium*. Cambridge: Harvard University Press.

Krishnamurthy, M. (2015) “We Can Make Rational Decisions to Have a Child: On the Grounds for Rejecting L.A. Paul’s Arguments”. En Hannan, S. y otros (eds.), P*ermissible Progeny?* Nueva York: Oxford University Press, 170–183.

Lackey, J. (2020), “Punishment and Transformation”. En Lambert, E. y J. Schwenkl (eds), *Becoming Someone New: Essays on Transformative Experience, Choice, and Change*. Oxford: Oxford University Press.

Laurence, S. y E. Margolis (1999).” Concepts and Cognitive Science”. En Laurence, S. y E. Margolis (eds.), *Concepts: Core Readings*. Cambridge: MIT Press, 3-81.

Lewis, D. (1989).” Dispositional Theories of Value”. *Aristotelian Society Supplementary Volume* 63 (1), 89-17.

Lee, A. (2023). “Knowing what it's like”. *Philosophical Perspectives* 37 (1), 187-209.

Lee, W. (2024). “What is Structural Rationality?”. *Philosophical Quarterly* 74 (2), 614-636.

Lifton, R. (1986). *The Nazi Doctors: Medical Killing and the Psychology of Genocide*. Nueva York: Basic Books

Lillehammer, H. (2000). “Revisionary dispositionalism and practical reason”. *The Journal of Ethics* 4 (3):173-190.

List, C. (2022) "Social Choice Theory". En Zalta, E. (ed), *The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/win2022/entries/social-choice,* ,consultado el 10 de julio del 2024.

Livingstone Smith, D. (2003). “Commentary on “On the Nature of Repressed Contents”. *Neuropsychoanalysis,* 5(2), 147–151.

Loar, B. (1990). “Phenomenal states”. *Philosophical Perspectives* 4, 81-108.

Lord, E. y D. Plunkett (2017). “Reasons Internalism”. En McPherson, T. y D. Plunkett (eds.), *The Routledge Handbook of Metaethics*. Nueva York: Routledge, 324-339.

Machery, E. (2009). *Doing without concepts*. New York: Oxford University Press.

Machery, E. (2010).” The bleak implications of moral psychology”. *Neuroethics* 3 (3), 223-231.

Manne, K. (2014).” Internalism about reasons: sad but true?”. *Philosophical Studies* 167 (1), 89-117.

Medvedev, D. y otros (2024)” The motivating effect of monetary over psychological incentives is stronger in WEIRD cultures”. *Nature Human Behaviour* 8, 456–470.

Mele, A. (2001). *Self-Deception Unmasked*. Princeton: Princeton University Press.

Mele, A. (2010). “Approaching Self-Deception: How Robert Audi and I Part Company,” *Consciousness and Cognition* 19,745–750.

Mele, A. (2019) “Self-Deception and Selectivity”. *Philosophical Studies,* 177, 2697–2711.

Millgram, E. (2000). “Mill’s proof of the principle of utility”. *Ethics* 110 (2), 282-310.

Mischel, W. (2009). “From Personality and Assessment (1968) to Personality Science, (2009)”. *Journal of Research in Personality* 43, 282–290.

Morris, L. S. y otros (2022). “On what motivates us: a detailed review of intrinsic *v.* extrinsic motivation”. *Psychological medicine*, 52(10), 1801–1816.

Nida-Rumelin, M. (1995). “What Mary couldn’t know: Belief about phenomenal states”. En Metzinger, T. (ed.), *Conscious Experience*, 219-41.

Nagel, T. (1970). *The possibility of altruism*. Oxford, Clarendon Press.

O’Dea, J. (2002). “The indexical nature of sensory concepts”. *Philosophical Papers* 32 (2), 169-181.

Papineau, D. (2007). “Kipke’s proof is ad hominem not two-dimensional”. *Philosophical Perspectives* 21, 475–494

Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Parfit, D. (1997). “Reasons and motivation”. *Aristotelian Society Supplementary Volume* 71 (1), 99–130.

Parfit, D. (2011). *On What Matters: Two-Volume Set*. Nueva York: Oxford University Press.

Paul, L.A. (2014). *Transformative Experiences*. Oxford: Oxford University Press.

Paul, L.A. (2015a). “Précis of Transformative Experience”. *Philosophy and Phenomenological Research* 91 (3), 760-765.

Paul, L. A. (2015b). “What You Can’t Expect When You’re Expecting”. *Res Philosophica* 92 (2), 1-23.

Paul, L. A. y J. Quiggin (2018). “Real-world problems”. *Episteme* 15 (3), 363-382.

Paul, L. A. (2020) “Who will I Become?” En Schwenkler, M. y E. Lambert (eds), *Becoming Someone New: Essays on Transformative Experience, Choice, and Change*. Oxford: Oxford University Press, 16-36.

Paul, L. A. y F. Cushman (2022). “Are desires interdependent?” En Doris, J. (ed), *The Oxford Handbook of Moral Psychology*. Oxford: Oxford University Press.

Perry, J. (2001). *Knowledge, Possibility, and Consciousness*. Cambridge: MIT Press.

Phelan, M. y W. Buckwalter (2012). “Analytic Functionalism and Mental State Attribution”. *Philosophical Topics* 40 (2), 129-154.

Prinz, J. (2007). “Mental pointing: Phenomenal knowledge without concepts”. *Journal of Consciousness Studies* 14 (9-10), 184-211.

Quine, W.V. O (1966). *The Ways of Paradox*, Nueva York: Random House.

Railton, P. (1986). “Facts and Values”. *Philosophical Topics* 14 (2), 5-31.

Railton, P. (2009). “Practical competence and fluent agency”. En Sobel, D. y W. Steven (eds.)*, Reasons for Action.* Nueva York: Cambridge University Press, 81-115.

Ranalli, C. y T. Lagewaard (2022). “Deep Disagreement (Part 1): Theories of Deep Disagreement”. *Philosophy Compass*, 17(12), 1-18.

Raz, J. (2009). “Reasons: Explanatory and normative”. En Sandis, C. (ed.)*, New Essays on the Explanation of Action*. Londres: Palgrave-Macmillan, 13-35.

Reisner, A. (2018). “Pragmatic Reasons for Belief”. En Star, D. (ed.), *The Oxford Handbook of Reasons and Normativity*. Nueva York, Oxford University Press, 756-782.

Ripstein, A. (2001). “Preference”. En Morris C.W. y A. Ripstein (eds), *Practical Rationality and Preference: Essays for David Gauthier*. Cambridge: Cambridge University Press, 37-55.

Rorty, A. (1988). “The deceptive self: Liars, layers, and lairs”. En. McLaughlin, B. y A. Rorty (eds.), *Perspectives on self-deception*. Berkeley: University of California Press, pp. 11-28.

Rosati, C. (1996).” Internalism and the good for a person”. *Ethics* 106 (2), 297-326.

Rothman, J. (2013,Abril). “The Impossible Decision”. *The New Yorker*. https://www.newyorker.com/books/page-turner/the-impossible-decision, consultado el 10 de julio del 2024.

Russell, B. (1911). “Knowledge by acquaintance and knowledge by description”. *Proceedings of the Aristotelian Society* 11, 108-28.

Sampson, E. (2021). “What if ideal advice conflicts? A dilemma for idealizing accounts of normative practical reasons”. *Philosophical Studies* 179 (4), 1091-1111.

Samuel, J. (2023). “Alienation and the Metaphysics of Normativity: On the Quality of Our Relations with the World”. *Journal of Ethics and Social Philosophy* 26 (1), 158-191.

Scanlon, T. M. (2004).” Reasons: A Puzzling Duality?”. En Wallace, R. J. (ed.), *Reason and value: themes from the moral philosophy of Joseph Raz*. Nueva York: Oxford University Press, 231-246.

Scanlon, T. (2014). *Being Realistic About Reasons*. Oxford: Oxford University Press.

Schroeder, M. (2007). *Slaves of the passions*. Nueva York: Oxford University Press.

Schroeder, M. (2014). *Explaining the Reasons We Share: Explanation and Expression in Ethics, vol. 1*. Oxford: Oxford University Press.

Scott-Kakures, D. (1996). “Self-deception and internal irrationality”. *Philosophy and Phenomenological Research* 56 (1), 31-56.

Shafer-Landau, R. (2003). *Moral realism: a defence*. Nueva York: Oxford University Press.

Singer, P. (2012). “The Objectivity of Ethics and the Unity of Practical Reason”. *Ethics* 123 (1), 9-31.

Sinnott-Armstrong, W. (1988). *Moral dilemmas*. Nueva York: Blackwell.

Sinnott-Armstrong, W. (2008). “Framing moral intuitions” En Sinnott-Armstrong, W. (ed.), *Moral psychology, Vol. 2: The cognitive science of morality: Intuition and diversity*. Cambridge: MIT Press, 47–76

Sinhababu, N. (2009). “The Humean Theory of Motivation Reformulated and Defended”. *Philosophical Review* 118 (4), 465-500.

Skorupski, J. (2010). *The domain of reasons*. Oxford: Oxford University Press.

Sobel, D. (2005). "Pain for objectivists: The case of matters of mere taste”. *Ethical Theory and Moral Practice* 8 (4), 437-457.

Sobel, D. (2009). “Subjectivism and idealization”. *Ethics* 119 (2), 336-352.

Smith, M. (1994). *The moral problem*. Cambridge: Blackwell.

Stalnaker, R. (1984). *Inquiry*. Cambridge: MIT Press.

Stanford, Kyle (2023). "Underdetermination of Scientific Theory”. En Zalta, E. (ed), *The Stanford Encyclopedia of Philosophy.* https://plato.stanford.edu/archives/sum2023/entries/scientific-underdetermination, consultado el 10 de julio del 2024.

Stratton-Lake, P. (2018). “Reasons Fundamentalism and Value”. En Star, D. (ed.), *The Oxford Handbook of Reasons and Normativity*. Oxford: Oxford University Press, 275-296.

Street, S. (2009). “In defense of future Tuesday indifference: Ideally coherent eccentrics and the contingency of what matters”. *Philosophical Issues* 19 (1), 273-298.

Sripada, C. (2014), “How is Willpower Possible? The Puzzle of Synchronic Self-Control and the Divided Mind”. *Noûs* 48, 41-74.

Sripada, C. (2015), “Moral Responsibility, Reasons, and the Self”. En Shoemaker, D. (ed.), *Oxford Studies in Agency and Responsibility* 3, 242-264.

Sundström, P. (2011). “Phenomenal Concepts”. *Philosophy Compass* 6 (4), 267-281.

Tye, M. (2003).” A theory of phenomenal concepts”. En O'Hear, A. (ed.), *Royal Institute of Philosophy Supplement*. Cambridge: Cambridge University Press, 91-105.

van Inwagen, P. (1996). “It Is Wrong, Everywhere, Always, for Anyone, to Believe Anything upon Insufficient Evidence”. En Jordan, J. y D. Howard-Snyder (eds.), *Faith, Freedom and Rationality*. Savage: Rowman and Littlefield, 137-154.

Wallace, R. J. (2020), "Practical Reason". En Zalta, E. (ed), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/fall2024/entries, consultado el 10 de julio del 2024.

Wedgwood, R. (2010). “The moral evil demons”. En Feldman, R. y T. Warfield (eds.), *Disagreement*. Oxford: Oxford University Press, 216-246.

White, R. (2009). “On Treating Oneself and Others as Thermometers”. *Episteme* 6 (3), 33-250.

White, N. M. (2011). “Reward: What Is It? How Can It Be Inferred from Behavior?” En Gottfried, J. A. (ed), *Neurobiology of Sensation and Reward.* Boca Raton: Taylor and Francis, 42-58.

Wiland, E. (2018). “Psychologism and Anti-psychologism about Motivating Reasons”. En Star, D. (ed.), *The Oxford Handbook of Reasons and Normativity*. Nueva York, Oxford University Press. 197-213.

Williams, B. (1979).” Internal and External Reasons”. En Ross H. (ed.), *Rational action: studies in philosophy and social science*. Nueva York: Cambridge University Press, 101-113.

Williamson, T. (2024). “A risky challenge for intransitive preferences”. *Noûs* 58, 360-385.

Wong, D. (2006). “Moral Reasons: Internal and External”. *Philosophy and Phenomenological Research* 72 (3), 536 - 558.

Worsnip, A. (2014). “Disagreement about Disagreement? What Disagreement about Disagreement?” *Philosophers' Imprint* 14 (18), 1-20.

Worsnip, A. (2018). “What is (In)coherence?” *Oxford Studies in Metaethics* 13, 184-206.