

Fallar es humano

un análisis filosófico de la falibilidad

Axel Arturo Barceló Aspeitia
abarcelo@filosoficas.unam.mx

I. Introducción

“No estoy convencido de que haya algo más que una etiqueta de moda [“normatividad”] que unifique de manera significativa esta gama de fenómenos que incluye la adhesión a los principios de la deducción, la toma de decisiones sobre cómo actuar, el adoptar creencias coherentes y útiles, el uso de palabras de acuerdo con sus significados, el pensamiento conceptual, el actuar de acuerdo a convenciones sociales, y el actuar como uno debe.”

Adam Morton (2006)¹

El objetivo de este libro es analizar la estructura de la falibilidad en sus diferentes manifestaciones normativas, especialmente en el ámbito de la epistemología y la ética. En particular, buscaré construir lo que podríamos llamar una teoría general de la falibilidad, delineando las relaciones entre una serie de conceptos normativos fundamentales como los de suerte, consecuencia, obligación, responsabilidad, preparación, negligencia, y similares. Trabajaré bajo la hipótesis de que esta estructura normativa es la misma en todos los campos normativos y lo ilustraré con ejemplos provenientes de los dos campos ya mencionados y otros más: del lenguaje, la estética, las ciencias cognitivas, etc. Si tengo éxito, la teoría que presento servirá para dar cuenta de fenómenos

¹. “I am not convinced that there is anything more than a fashionable label in common [– “normativity” –] that makes a significant unity of [this] wide range [of phenomena] that includes adhering to principles of deduction, choosing actions, making beliefs coherent and serviceable, using words in accordance with their meanings, thinking conceptually, following social conventions, and acting as one ought.” Traducción mía.

tan aparentemente disímbolos como: cómo prepararse de manera responsable para un desastre natural hasta porqué es tentador pensar que el éxito es relativo a quién y desde qué circunstancias se evalúa, pasando por cómo resolver la paradoja del *sorites*, y porqué es posible que la gente no se ría al escuchar un chiste gracioso. Si mi teoría es correcta, todos estos fenómenos son manifestaciones de la falibilidad.

Lo que no haré será ofrecer una explicación de la normatividad en términos no normativos. Dudo que la normatividad no sea irreducible, pero esa no será una de las cuestiones que abordaré. Otro monstruo que dejaré dormir a mi lado será el del escepticismo. En otras palabras, no diré nada sobre cómo “responder al reto escéptico”, lo que sea que eso signifique. Mas bien, trabajaré bajo el supuesto de que el discurso normativo tiene sentido, es decir, que no da lo mismo actuar de manera responsable que irresponsable, que el éxito y el fracaso son ambos posibles y que hay una diferencia genuina entre hacer las cosas bien y hacerlas mal. Muchos filósofos han presentado argumentos muy brillantes contra de estos supuestos de sentido común, pero no los trataré de refutar aquí.

A fin de cuentas, defenderé un dualismo contextualista y materialista. Mi propuesta es un contextualismo en tanto sostiene que por lo menos algunas de las nociones centrales de las que dependen nuestras evaluaciones normativas son sustancialmente sensibles al contexto. En particular, qué posibilidades sean previsibles y a cuales de ellas sea necesarios atender para actuar de manera responsable en una circunstancia dada dependerá de manera sustancial en qué recursos tengamos disponibles, tanto en el contexto en el que actuamos, como en el contexto desde el cual evaluamos nuestra acción. Por ello, mi propuesta también es materialista (en el sentido presente en el trabajo de filosofas como Sally Haslanger (2000), Toril Moi (1999) y en mi trabajo previo con Ángeles Eraña (2016), es decir, en el sentido de que las circunstancias materiales desde las cuales actuamos y evaluamos son relevantes par nuestros juicios normativos, no en el sentido metafísico de rechazar la existencia de entidades no materiales).

Mi propuesta es un dualismo en tanto sostiene que, respecto a la pregunta de *¿por qué preferimos un logro genuino, es decir, uno que emana de nuestros esfuerzos por alcanzarlo, sobre uno que se obtiene mas bien por suerte?*, tanto deontologistas como consecuencialistas tienen razón, pero cada uno de su dominio normativo propio e independiente. Del deontologismo recupero la tesis de que no es posible dar cuenta de todas nuestras prácticas

normativas apelando sólo a nuestros fines, (y a nociones derivadas de ellas como las de confiabilidad, probabilidad, robustez, etc.) como sostienen los consecuencialistas. Sin embargo, mi dualismo recupera también las intuiciones consecuencialistas de que sería raro evaluar positivamente medios que no nos acercaran a nuestros fines (aunque también creo que aunque raros, estos casos sí existen, como sostienen los que llamaré deontologistas radicales). Esto significa, en el caso moral, que es razonable esperar de un agente virtuoso que actué bien y que sus actos no tengan consecuencias dañinas de la misma manera que es razonable esperar de quién no nos hace daño y actúa de manera impecable que sea una persona de carácter virtuoso, aun cuando el tener un carácter virtuoso no pueda reducirse a la tendencia confiable de actuar bien y no dañar a otros (y, de manera inversa, el actuar bien no pueda reducirse al actuar virtuoso). El que ambos aspectos normativos sean diferentes, aunque relacionados es lo que caracteriza mi postura dualista.

Sin embargo, también es uno de mis intereses centrales en este libro el presentar la estructura general de nuestras prácticas normativas. Por ello, quiero dejar claro qué aspectos aceptan (o, aun mas, deberían aceptar) tanto deontologistas como consecuencialistas (y otros que adoptan posturas derivadas de ellas, como los relativistas, gradualistas, etc.). Por ello, al mismo tiempo que presento mi propia propuesta, señalaré los puntos y las direcciones en las cuales podrían desarrollarse propuestas alternativas. En otras palabras, trato de ser justo con las motivaciones y virtudes de otros tipos de teorías de la normatividad. Hago esto no sólo por razones de ética profesional – es mas honesto reconocer las fortalezas y razones de las propuestas alternativas que atacar muñecos de paja – sino también por razones propedeúicas. He podido llegar a mi propuesta principalmente porque las propuestas consecuencialistas y deontologistas actuales han alcanzado un grado de sofisticación superlativo. En otras palabras, aun cuando termino creyendo que tanto el deontologismo como el consecuencialismo están equivocados, creo que cada uno de ellos acierta a resaltar aspectos fundamentales de nuestras prácticas normativas que serían difíciles de apreciar de otra manera.

La teoría estará estructurado alrededor de *cinco* conceptos centrales del fenómeno normativo: falibilidad, suerte, fin, medio y consecuencia. En epistemología, se dice que el conocimiento es falible en tanto es posible que un agente conozca algo en base a una justificación que no garantiza la verdad de aquello que conoce. En base a esta caracterización, definiré un proceso como falible, en general, si su éxito no está garantizado, es decir, si el

llevar a cabo el proceso de manera correcta no garantiza que se alcance su objetivo. El objetivo de este libro es explorar la estructura normativa de este tipo de procesos. Hoy en día, en filosofía, cuando se habla de falibilismo, se piensa automáticamente en la tesis epistemológica según la cual ninguna creencia humana está justificada de manera infalible. Sin embargo, la falibilidad epistémica es sólo un caso de un fenómeno más general. Que algo sea falible no significa más que puede fallar y que algo sea infalible, por lo tanto, significa que garantiza su objetivo de manera concluyente. Así pues cualquier actividad que tenga un fin u objetivo, puede ser falible – si es posible que no alcance su objetivo – o infalible – si lo garantiza.

Los procesos falibles generan un fenómeno muy interesante: dado que no garantizan su éxito, permiten que sea posible alcanzar el éxito, no en virtud de haber llevado a cabo el proceso de manera correcta, sino por mera suerte. En este libro estudiaré este fenómeno tal y como se manifiesta en diferentes espacios normativos, una vez más, poniendo especial atención al ámbito de lo epistemológico y de lo ético. En particular, identificaré dos preguntas importantes que han dominado el debate alrededor de la falibilidad: la pregunta descriptiva de ¿qué distingue un caso de éxito genuino de uno que se da solo por suerte? y la pregunta fundacional de ¿porqué preferimos los primeros sobre los otros?

Respecto a la primera pregunta, propondré distinguir diferentes tipos de normatividad – una dirigida hacia los fines, otra a los medios y, finalmente, una tercera para los logros. Pese a ser diferentes, estos tres tipos de normatividad, sostengo, están íntimamente ligados, es decir, un buen medio está dirigido a su fin y dicho fin se logra sólo cuando medio y fin están adecuadamente relacionados, es decir, cuando el fin se alcanza a través del medio que dirigimos hacia él y no por mera suerte. Sin embargo, aunque relacionados, también son relativamente independientes. Esto significa que un proceso falible que no resulta en aquello que persigue puede ser, sin embargo, correcto. Esto significa, en el caso epistémico, que hay algo que podemos evaluar positivamente en un caso de creencia falsa justificada o, en ética, que podemos evaluar como moralmente correcto un acto que se realizó de manera responsable pero tuvo consecuencias dañinas para una víctima inocente.

Compararé mi propuesta con otras propuestas en la literatura, especialmente con el deontologismo, el consecuencialismo y el relativismo. Según el consecuencialismo, lo realmente importante para evaluar un medio es si nos acerca a nuestro fin o no. Para el deontologista, en contraste, lo importante más bien es cómo

administramos nuestros recursos en persecución de dicho fin y no si, a fin de cuentas, lo alcanzamos o no. Finalmente, según el relativista, nuestra evaluación de un proceso falible varía conforme vamos teniendo mayor o menor certeza de que alcanzara su objetivo o no. En este sentido, lo correcto o incorrecto de un acto puede cambiar conforme va produciendo consecuencias positivas o negativas.

Finalmente, en el último capítulo subrayaré la importancia de los presupuestos implícitos bajo los cuales esperamos que un proceso realizado de manera irreprochable deba resultar en éxito y cómo se relacionan con lo que en derecho (y en lógica y epistemología también) se han llamado *derrotadores*. En general, defenderá la tesis de que cuando logramos algo, no logramos todas sus partes (en el sentido de todas sus condiciones necesarias) y, en particular, no logramos la presencia ni ausencia de derrotadores. La presencia o ausencia de derrotadores, por definición, nunca está en nuestras manos, aun cuando sean condición necesaria de nuestro logro o justificación.

Como cualquier proyecto de esta magnitud, esta obra no habría sido posible sin el apoyo de un gran número de colegas con quienes pude discutir muchas de las ideas que aquí expreso. Agradezco de manera especial los comentarios de Ángeles Eraña, Miguel Ángel Fernández, Catherine Elgin, Stephen Yablo, Mark Richards, John MacFarlane, Pedro Stepanenko, Ricardo Mena, Fernando Rudy, Sally Haslanger y Carlos Muñoz Suárez. Porciones de este libro fueron presentados en los Seminarios de Filosofía del Lenguaje y de Investigadores del Instituto de Investigaciones Filosóficas de la Universidad Nacional Autónoma de México y en el Encuentro UNAM-Universidad de Texas en Austin. Agradezco los comentarios recibidos. Mención especial merecen Nancy Abigaíl Nuñez con quién discutí largamente estos temas y cuyo trabajo de doctorado debería leerse como obra hermana a ésta, y los estudiantes del curso Temas Selectos de Filosofía del Lenguaje, con quienes revisé un borrador completo de este libro durante el primer semestre del año 2017: Adrián Ávila Mejía, Luis David Abeijon Malvaez, Arturo Abeijon Malvaez, Erick de la Mora, Sofía Ortega, Amalinali Armendáriz, Fabio Spada, Nitzhui Daniela Morales Pineda, Carlos Romero, Cristian Pineda, Angel Cantú, Cassandra Valdez, Rómulo Antonio Salazar Matus y Fernanda Montoya. Ellos se merecen mi mas profundo agradecimiento.

II. Falibilidad

La gente ... siempre se pregunta si sabías que iba a ser un fracaso. Dicen cosas como, "¿No lo sabían?" "¿Cómo podrían no haberlo sabido?" Mi experiencia es que no lo sabes. No lo sabes porque tienes mucho invertido, y un interés personal [y] porque tienes esperanza.²

Nora Ephron (2010, 109-110)

A. Infalibilismo y Falibilidad

La pregunta que guía esta reflexión es: ¿qué papel juega el mal, el error, la falla en un discurso o práctica normativa? Aquí veo yo dos opciones: verla como un enemigo con el que hay que luchar a muerte, o tratar de incorporarla como algo de lo que nuestra teoría debe tomar en cuenta, con lo que nuestra práctica debe lidiar. La primera opción resulta en un infalibilismo en el que lo que se busca es desterrar y evitar en lo absoluto el mal, el error, la falla, etc. La segunda opción resulta en un discurso y práctica normativa que le da a la falibilidad un lugar central. Como bien señala Dutant (2015), durante mucho tiempo, la filosofía occidental optó por la primera de estas alternativas, pero en años más recientes la falibilidad se ha instaurado como un aspecto ineludible de la normatividad. Lo que me interesa en este libro es estudiar precisamente las consecuencias de apelar a la falibilidad como manera de darle un lugar al mal, el error, la falla, el fracaso, la falsedad, el daño, etc. dentro de nuestro discurso y práctica normativa. Para el infabilista, lo que está mal, está mal y punto; no hay más. Para el no infabilista, sin embargo, hay sentidos y circunstancias en las cuales parece que **está bien estar mal**, lo cual, por supuesto, suena paradójico; el objetivo de este libro es mostrar como no hay nada realmente paradójico aquí.

Para entender mejor el contraste entre una normatividad infabilista y una que no lo es, pensé en posibles ejemplos. El más obvio es el epistemológico, donde el contraste se manifiesta entre una concepción de la epistemología en donde, para parafrasear el título completo del *Discurso del Método* de Descartes (1637), conducir

². People... always wonder if you knew it was going to be a flop. They say things like, "Didn't they know?" "How could they not have known?" My experience is that you don't know. You don't know because you are invested [and] because you hope. Mi traducción.

bien la razón no es mas que buscar la verdad. En otras palabras, el único objetivo es la verdad y lo que debé de hacer es aquello que – de manera infalible – nos lleve a ello. En contraste, hoy en día, se piensa que esto, o bien es imposible, o es insuficiente para dar cuenta de lo que llamamos “conocimiento” y su relación con la verdad. Pero no quiero quedarme en el ejemplo epistemológico, por un lado porque ha sido ya muy elaborado, tanto en el texto de Durant como en otros lados, sino porque no quiero dar la impresión de que el fenómeno es eminentemente epistemológico – aunque diré mas sobre esto mas adelante.³ Entonces pensé en cualquier otro mal. Por ejemplo, durante una parte importante del siglo veinte se pensó que la lógica cumplir con su función normativa simplemente con señalar cuándo un argumento o inferencia era válido, es decir, cuando su conclusión se seguía lógicamente de sus premisas. Esto significaba que, frente a casos de argumentación o inferencia inválida, la lógica no tenía otra cosa que aportar mas que decir eso, es decir, que son inválidos. En este sentido, su postura normativa era infalibilista. A finales del siglo pasado, varios filósofos y científicos cognitivos – provenientes predominantemente de la psicología cognitiva y de la inteligencia artificial, peor también de lo que ahora se llama la lógica informal –, propusieron adoptar una postura distinta que reconociera que no bastaba señalar que dichas inferencias o argumentos eran falaces, sino que era necesario ampliar el alcance del discurso normativo sobre dichos argumentos para hacer distinciones mas finas. En otras palabras, no aceptaban simplemente la aparente conclusión de que realizar dichas inferencias o apelar a dichos argumentos siempre era irracional, sino que a veces podría también ser racional – sin dejar de ser lógicamente inválidos. Nótese que lo que buscaban no era meramente explicar porqué a veces cometé errores al razonar o argumentar, es decir, su proyecto no era meramente descriptivo, sino eminentemente normativo. Su objetivo fue ampliar el discurso normativo sobre la argumentación y la inferencia para incorporar al error y distinguir entre errores racionales y otros genuinamente irracionales.

³. Aunque tampoco quiero dar la impresión contraria, es decir, que el fenómeno no es epistemológico, o que es eminentemente ético o económico o evolutivo, etc. Por eso no me gusta hablar, como Bortolotti (2016) de “La Ética de la Creencia” como algo distinto de la epistemología. El que todos estos regimenes normativos tengan la misma estructura no hace que todos sean epistemología, o que todos sean ética, etc.

Para entender mejor el paso de una filosofía infalibilista a una falibilista, vale la pena revisar también el seminal trabajo de John L. Austin y, en especial, su conferencia “Un alegato en pro de las excusas” (1956, 1957). En este seminal conferencia Austin señala que nuestras prácticas éticas no se reducen meramente a decir que ciertas cosas son buenas y otras malas, sino que son mucho más complejas: involucran justificaciones, excusas, perdón, etc. Y Austin nos invita a notar que lo único que se puede excusar es lo malo, es decir, los errores, las fallas. Las excusas tienen sentido precisamente porque no basta señalar al mal y decir que es malo. Hay que hilar mas fino y distinguir entre males justificados y no justificados, excusables e inexcusables, perdonables e imperdonables. Una vez más, la idea fundamental es que la normatividad no puede excluir al mal, a la falla, el error, sino que debe darle espacio (no como mero enemigo a evitar). En otras palabras, lo que señala Austin es que nuestras prácticas normativas morales no son infalibilistas, sino que reconocen nuestra falibilidad.

Desde una perspectiva infalibilista es difícil hacer sentido de nuestra práctica ordinario de pedir disculpas u ofrecer excusas. Si pensamos que simplemente hay cosas buenas o malas, el que ofrece disculpas o bien hizo algo bueno o algo malo. Si hizo algo bueno, no necesita disculparse (a lo mas necesitaría explicar porque lo que hizo es bueno); si hizo algo malo, tampoco tiene sentido que se disculpe, pues lo que hizo sigue siendo malo y nada que diga lo va a cambiar. De cualquier manera, no tendría sentido ofrecer disculpas. Sin embargo, aceptar que la práctica de pedir disculpas no tiene sentido es aceptar que una práctica normativa muy extendida no tiene sentido. Aceptar una conclusion tan fuerte debería requerir también evidencia muy fuerte. Es preferible buscar otra explicación que no nos presente a los humanos como sistemáticamente envueltos en una práctica sin sentido.

B. Una Nueva Metodología para la Filosofía

Austin es conocido como uno de los principales proponentes de lo que hasta la fecha se conoce como la “Filosofía del Lenguaje Ordinario”. Una de las tesis principales de este movimiento filosófico era darle un lugar central en la reflexión filosófica a la práctica lingüística habitual de los hablantes comunes y en ese momento se presentaba como revolucionaria en frente a la tendencia, dentro de la filosofía analítica, de ‘reformular’ al lenguaje de manera artificial para que mejor reflejará las relaciones conceptuales de interés científico y filosófico. En otras palabras, frente a quienes pensaban que el lenguaje, aunque una buena herramienta para la comunicación

ordinaria, era inapropiada para el análisis conceptual, Austin y sus colegas Ryle (1953) y Strawson (1959) proponían voltear a nuestro uso ordinario del lenguaje como guía, falible por supuesto, a esas relaciones conceptuales. Según Austin, ver porqué en nuestra práctica lingüística habitual, es decir, en como hablamos y escribimos en nuestras lenguas naturales, hacé ciertas distinciones – por ejemplo, entre “éxito” y “logro” – y relacionamos ciertas cosas – por ejemplo, todas las cosas para las que usamos el verbo “hacer” – nos permite ver distinciones y similitudes que de otra manera serían difíciles de identificar. Tal y como señala Austin:

“...nuestro surtido común de palabras encarna todas las distinciones que los hombres han considerado lo suficientemente importantes como para ser delineadas, y las conexiones más dignas de ser señaladas en la vida de muchas generaciones; éstas, seguramente, son más numerosas y sonoras desde que los hombres lograron imponerse contra la larga prueba de supervivencia del más apto; y más sutiles, por lo menos en los temas ordinarios y razonablemente prácticos, que los que tú o yo es posible pensar desde nuestro sillón en una tarde –el método alternativo más favorable.”

Este voltear a la manera como habitualmente usamos el lenguaje comparte con otro par de giros metodológicos en filosofía muy importantes del último siglo – el naturalismo y el giro pragmático – un interés por restituir el balance entre lo concreto y lo abstracto en filosofía. Llamamos “giro pragmático” a la tendencia, comúnmente asociada con el segundo Wittgenstein y el pragmatismo norteamericano, de estudiar los problemas filosóficos no ‘en abstracto’, sino en su uso práctico. Por ejemplo, en el caso de la normatividad, el consejo es dejar de preguntarnos qué es el bien, el mal, el deber, el valor, etc. en abstracto y empezar a pensar mas bien en cómo se manifiestan en lo que hacé, es decir, poner las prácticas de evaluación al centro de la reflexión filosófica sobre la normatividad. Por el otro lado, el naturalismo puede caracterizarse como una reacción en contra de las excesivas idealizaciones que guiaron al análisis conceptual de principios del siglo XX y a favor de un renovado interés en voltear a ver las cosas en toda su complejidad. Para nuestros objetivos, adoptar una perspectiva naturalizada sobre las prácticas normativas significa reducir el número de idealizaciones que asumimos y voltear, de manera empírica, a nuestras prácticas normativas ordinarias, es decir, a cómo de hecho evaluamos. Al combinar el

naturalismo con el giro pragmático, se abre una brecha metodológica por la cual es posible dejar a nuestras prácticas evaluativas ordinarias guiarnos hacia un mejor entendimiento de la normatividad.⁴

Y aquí es muy importante señalar que cuando hablamos de nuestras prácticas evaluativas nos referimos a la manera en la que hacemos juicios evaluativos nosotros, aquí y ahora. Esto nos obliga también a voltear la cara hacia las disciplinas que mejor nos pueden servir para entender quiénes somos estos nosotros aquí y ahora que hacemos dichos juicios evaluativos. El seminal trabajo de John L. Austin aparece también aquí como un buen ejemplo. En él Austin ya apunta hacia dos fuentes importantes de información para pensar la normatividad: la psicología y las leyes. A lo largo de este texto yo también beberé de ambas fuentes. De la psicología será fundamental reconocer la importancia de los sesgos epistémicos sistemáticos de nuestras capacidades cognitivas y del derecho la importancia del concepto de negligencia o descuido para asignar responsabilidad a agentes por las consecuencias negativas de sus actos. Pero también podríamos añadir a la lista otras disciplinas como la economía o la antropología.

Podríamos pensar a la psicología como la disciplina que nos ayudará a situar la evaluación dentro del marco general de la vida mental del sujeto y al derecho y la antropología como las que nos ayudarán a situarla en su marco social más amplio. A fin de cuentas, la evaluación es algo que comúnmente hacemos los individuos, sobre individuos, pero también como colectividades y sobre colectividades; a veces en el foro interno de nuestras mentes y otras veces en el foro público. De ahí la relevancia de tomar en cuenta tanto lo que el derecho tiene que decir, como la psicología o la antropología.

De la psicología del razonamiento, es muy útil comprender el fenómeno de los sesgos epistémicos. Tal como lo define Claudia Lorena García (2004), “un procedimiento o mecanismo cognitivo de un determinado sujeto” muestra un sesgo epistémico cuando “hay un rango de condiciones o contextos en los cuales el procedimiento o mecanismo ... produciría sistemáticamente resultados cognitivos que serían epistémicamente

⁴. Adoptar esta metodología, por supuesto, es una decisión controversial. En conversación personal, por ejemplo, Catherine Elgin me decía que reconocía que su consecuencialismo, como veré más adelante, no da cuenta de cómo de hecho evaluamos las acciones de otros; sin embargo, para ella, la filosofía debía también decirnos de las cosas que hacemos, cuales no deberíamos.

inaceptables o injustificados, y otro rango de condiciones o contextos en donde el mecanismo en cuestión produciría en general resultados epistémicamente aceptables o justificados.” (García-Aguilar 2004, 82) Dada la sistematicidad de dichos errores, es difícil pensar que se deben a meros errores de ejecución (Eraña 2003), es decir, que nuestra mente tenga la capacidad de razonar siempre de manera lógica impecable y que, sin embargo, algunas veces, por distracción o descuido, nos equivocamos. Mas bien, lo que parece suceder es que nuestras capacidades cognitivas no están diseñadas para producir siempre dichos resultados lógicos. Es pues, necesario preguntarnos el porqué de dichos errores, es decir, si no es para producir siempre resultados lógicamente impecables, cuál es la función de nuestros mecanismos cognitivos. En otras palabras, es necesario, insisto, reintegrar el error lógico en nuestra teoría normativa del razonamiento y la cognición. No es posible simplemente decir que somos ilógicos. Tal y como lo explica Ángeles Eraña:

“A partir de la evidencia (empírica), los psicólogos cognoscitivos afirman que los seres humanos llevamos a cabo inferencia a través de una serie de atajos también llamados *reglas* o *procesos heurísticos* que se caracterizan por mostrar algunos sesgos en nuestro razonamiento [resultantes de ignorar aspectos relevantes del problema, o prestarle demasiada importancia a aspectos irrelevantes, que] deben atribuirse a la utilización de dichas reglas, esto es, a la utilización de sistemas de instrucciones que nos permiten solucionar un problema, no de manera exacta ni en todos los casos pero sí de manera que no [requiera] una computación excesiva desde el punto de vista de los recursos cognitivos que tenemos disponibles” (Eraña 2003)

Este mismo contraste aparecerá una y otra vez cuando consideré ejemplos en otros campos normativos: el contraste entre un método (herramienta, procedimiento, mecanismo, acto, método, práctica, etc.) infalible pero muy caro (o hasta inaccesible) en términos de recursos (de tiempo, atención, esfuerzo, conocimiento, control, etc.) y otro más accesible (en el sentido de requerir de menos recursos para implementarse) pero falible. Si pensamos que la única manera de juzgar un método es preguntarnos si nos da los resultados correctos o no, entonces los métodos de primer tipo serían los únicos métodos “buenos”, mientras que una noción mas amplia de la normatividad podría explicar porqué es preferible optar por métodos del segundo tipo, los cuales, aun cuando no siempre den los resultados correctos, siguen siendo “buenos” en algún sentido.

Tratar de domesticar este tipo de errores no significa que no se den de hecho errores de ejecución, es decir, que no cometamos errores por descuido sino que todos nuestros errores sean, en el fondo, racionales. Por supuesto que existe la negligencia, pero no todos los errores se deben a ella. Algunos errores se deben a la mala suerte, por ejemplo. Es por ello que en derecho se hacen tantas distinciones, entre crímenes y delitos, entre descuido y malicia, etc. No todos los errores son iguales y por eso no los tratamos igual. Algunos los dejamos ir, otros los señalamos; algunos se pueden perdonar tras una disculpa, mientras que otros son imperdonables; algunos involucran la suerte, otros la falta de preparación, etc. Una teoría adecuada de la normatividad, por lo tanto, debería de ser capaz de dar cuenta de esto y una teoría infabilista claramente no puede. En el resto del libro presentaré una teoría general de la normatividad cosntruída ex-profeso para lidiar con este tipo de fenómenos, es decir, una teoría de la normatividad que le da un lugar central a la falibilidad.

De la economía tomamos la idea de que, una vez que nos proponé algo, no hacé todo lo que es posible hasta garantizar lograrlo. Si quiero un teléfono celular nuevo, no voy a pagar lo que sea por él, aun cuando cuente con el dinero. Si mi capital fuera infinito, sin lugar a dudas tal vez pagaría lo que fuera por él; pero no es así, mis recursos son limitados y como tales debo administrarlos. Esto significa determinar hasta cuanto sería responsable pagar por dicho teléfono, dadas mis otras necesidades, deseos, etc. En este libro no nos detendré a estudiar todos los factores que afectan nuestra decisión de dedicarle ciertos recursos a ciertos fines, y no a otros. Lo único que es importante para nuestros objetivos es reconocer que nuestras decisiones y juicios dependen de nuestras capacidades y recursos limitados. Que el que no hagamos todo lo que está en nuestras manos para conseguir un fin no significa que no lo valoramos o que, en realidad, no es uno de nuestros fines.

Para ilustrar la importancia filosófica de esta idea volvamos al ejemplo epistemológico y pensé en una de las críticas medulares que, desde el pragmatismo (Peirce 1868), se ha hecho al infalibilismo de Descartes (1641): si tuviéramos recursos y capacidades ilimitados, tal vez un fundacionismo como el de Descartes tendría sentido, es decir, tendría sentido ir construyendo nuestro conocimiento a partir de cero basándonos sólo en certezas inamovibles que nos garanticen por completo la verdad de lo que creé. Sin embargo, no somos agentes epistémico perfectos con recursos ilimitados, sino seres finitos que, además de buscar el conocimiento, tenemos otros fines en la vida; y a veces, para alcanzar estos otros fines – es decir, para vivir en este mundo – necesitamos

tomar decisiones restringidas en el tiempo y con información escasa. Es decir, necesitamos decidir qué creer en estados de incertidumbre. Para dar cuenta de estas decisiones, para dar cuenta de estas creencias, una epistemología infalibilista no está bien preparada.

De manera similar, una concepción infalibilista de la racionalidad humana tampoco puede dar cuenta de las reglas heurísticas que de hecho subyacen a nuestra competencia cognitiva. Las adoptamos porque nos permiten ahorrar recursos, pero ello no significa que no valore la corrección lógica o epistémica, sólo significa que no estamos dispuestos a perseguirla a cualquier costo. Dado que tenemos otras necesidades y que vivimos con ciertas limitaciones, es útil contar con capacidades cognitivas de este tipo.

C. Metas y Medios

“El hombre que pretende verlo todo con claridad antes de decidir nunca decide.” Henry-Frédéric Amiel

Hace unos meses, tuve la oportunidad de escuchar a Miriam Schoenfeld decir algo muy obvio, pero no por ello poco profundo, sobre la epistemología. Decía que, si uno valora la verdad, para la pregunta ¿qué debo creer? había una respuesta simple y obvia: debo creer lo que es verdad. Sin embargo, si la pregunta de ¿qué debo creer? se respondiera tan fácilmente, los libros de epistemología serían mucho, mucho más cortos. Para entender porqué, permítaseme poner otro ejemplo completamente análogo pero más obvio y hasta gracioso. Supongamos que alguien presumiera saber la manera de hacerse rico en la lotería, es decir, qué debo hacer para hacernos ricos jugando la lotería y le preguntaré cuál es esa manera. Supongamos entonces que su respuesta fuera “apostar siempre al número ganador”. En sentido estricto, lo que nos dice es cierto. Apostar siempre al número ganador es una manera infalible de hacerse rico jugando a la lotería. Efectivamente, si apostamos siempre al número ganador, haré mucho dinero en la lotería. Sin embargo, es claro que eso no es lo que pedimos cuando pedimos por la manera de hacernos ricos en la lotería. Lo que pedimos es un *cómo*, es decir, algo como un método que nos guíe en qué hacer para ganar la lotería. Apostar al billete ganador es, más bien, una definición de *lo que es* ganar la lotería. En otras palabras, lo que pedimos fue un medio, y lo que se nos dio fue (una redescipción de) la meta.

Lo mismo sucede con la pregunta ¿qué debé creer? Responder que debé creer la verdad, aun cuando es cierto, no nos da lo que pedimos; y esto es así porque no estamos meramente buscando una definición de creencia *buena*, una propiedad que distinga aquello que debé creer de lo que no. Al igual que en el caso de la lotería, lo pedimos un *qué* sino un *cómo*: un método, un mecanismo, una manera de alcanzar nuestra meta. No pedimos una re-descripción de nuestras metas, lo que pedimos son los medios para alcanzarlas. Pedimos algo que nos permita guiar nuestro comportamiento y deliberación. Necesitamos una caracterización de qué hacer, qué creer, qué decidir, etc. que nos permita, a partir de nuestras capacidades y recursos reales, tomar decisiones, adquirir creencias, etc. apropiadas.

Comparé dos posibles reglas epistémicas: “debé creer cosas verdaderas”, “debé creer sólo aquello de lo cual no es posible dudar”. Aun en una teoría epistémica en la que ambas fueran verdaderas, estos dos tipos de reglas deben jugar papeles distintos. Como señalé en el párrafo anterior, el primer enunciado puede servir para definir una de nuestras metas epistémicas, pero no sirve para guiar nuestra deliberación epistémica; esto se debe a que para aplicarla, deberíamos saber algo que no sabemos a la hora de la deliberación. Al igual que la regla “para ganar, hay que apostar el número ganador”, es inservible como estrategia o método. Cuando lo que pedimos es una regla que nos pueda guiar en la acción, deliberación, etc. necesitamos reglas como la segunda, “debé creer sólo aquello de lo cual no es posible dudar”, pues para aplicarla sólo necesitamos apelar a cosas que sí están a nuestra disposición, como nuestros propios estados mentales conscientes, que son accesibles a nuestra reflexión. En otras palabras, al momento de la deliberación, necesitamos reglas que apelen sólo a recursos y capacidades que tenemos de hecho disponibles, y no a cosas que trascienden nuestras capacidades y recursos.



Con esta idea en mente, permítaseme introducir una distinción entre dos maneras de hablar de lo que debé de hacer (o, en general, de usar lenguaje normativo). Llamaré **trascendentes** a enunciados normativos como el ya mencionado “debé creer cosas verdaderas”, para contrastarlos con enunciados normativos **instrumentales** como, por ejemplo, “debé creer sólo aquello de lo cual no es posible dudar”. Aun cuando ambos tipos de enunciados nos dicen qué debé de hacer, las diferencias entre ellos son sustanciales. Si bien los

primeros nos sirven para establecer las **metas** de nuestro actuar, no sirven de mucho a la hora de determinar los **medios** para alcanzar dichos fines. Esto se debe a que el vocabulario que se usa para describir dichas metas refiere a aspectos de la realidad que trascienden nuestras capacidades y/o están fuera de nuestro control – de ahí que haya decidido llamarlas “trascendentes”. La noción de ‘verdad’ nos trasciende de una manera que la de ‘duda’ no lo hace. Que una creencia que tengamos sea verdadera o no depende de manera sustancial en cosas que no están bajo nuestro control o requiere para su reconocimiento recursos o capacidades de las que no disponé, mientras que la duda, en tanto carácter subjetivo de nuestro estado mental, es de fácil acceso directo, es decir, nos es posible dar fácilmente dar cuenta de ella.

Nótese que la distinción se aplica no sólo en el campo de la epistemología. Cuando juzgamos moralmente un acto por sus consecuencias, por ejemplo, usamos un criterio trascendente, pues las consecuencias de nuestros actos no están por completo bajo nuestro control ni tenemos los medios disponibles para preverlas por completo. Sin embargo, cuando las evaluamos por sus intenciones, o evaluamos el carácter de alguien por sus motivos, usamos criterios instrumentales, pues asumimos que nuestras intenciones y motivos sí dependen de nosotros de una manera en que las consecuencias de nuestros actos no (Williams 1981). En general, la distinción tiene sentido cada vez que nos interesa algo que no está completamente a nuestro alcance y aun así, tratamos de conseguirlo. Además de epistemología y moral el mismo fenómeno es conocido también en estética, por mencionar otra área central de la filosofía, donde el artista muchas veces busca provocar una reacción en su público, pero esta reacción no depende completamente de ella. Así, por ejemplo, en comedia suele distinguirse entre el hacer reír y el ser gracioso, el primero es el fin trascendental y el segundo el instrumental. Aunque relacionados, son valores distintos y se puede dar uno sin el otro: algo puede ser gracioso y no hacer reír a todos y viceversa, algo puede hacer reír a alguien sin ser gracioso y hasta algo puede ser gracioso, hacer reír a alguien y, sin embargo, no hacerlo reír **por** ser gracioso, sino por otra razón. La misma distinción existe entre divertir y ser divertido, hacer llorar y ser triste, etc.

En investigación científica, por poner otro ejemplo, nos interesa la investigación que da resultados valiosos, pero desafortunadamente no es posible saber de antemano que investigación nos dará este tipo de resultados, por eso he desarrollado otro tipo de criterios para evaluar qué investigación desarrollar, como por

ejemplo, la eminencia de los investigadores que la proponen. El primero es un valor trascendente; el segundo, un valor instrumental. Aun los investigadores más eminentes pueden hacer mala investigación y, viceversa, investigadores completamente desconocidos pueden estar desarrollando investigación de excelencia, además de que la eminencia científica responde a muchos criterios además de la calidad de la investigación que se realiza (Shiota, *manuscrito*); sin embargo, tal vez de manera errónea, confiamos en que la eminencia nos sirva de guía falible en nuestra búsqueda de investigación de calidad.

Esta distinción corresponde *a grosso modo* a la vieja distinción entre dos tipos de reglas que identificó Rawls en (1955): normas que seguimos al interior de una práctica y valores que justifican dicha práctica, aunque Rawls mismo reconoce que la distinción aparece ya en la obra de Hume, Mill, Austin, Toulmin, y otros. La distinción queda claramente ejemplificada en la famosa anécdota apócrifa sobre el ladrón de bancos Slick Willie Sutton que apareció originalmente en *The Saturday Evening Post* en enero de 1951:

Una vez, alguien le preguntó a Slick Willie Sutton, el ladrón de bancos, por qué robaba bancos.

La pregunta podría haber revelado una historia de injusticia y búsqueda permanente de venganza. A lo mejor un banquero había embargado la granja de su familia, o tal vez la hija de un banquero había despreciado a Sutton por otro hombre.

Sutton, quedó sorprendido, como si le hubieran preguntado "¿Por qué un fumador enciende un cigarrillo?"

"Robo bancos porque ahí es donde está el dinero", dijo...(Yoder 1951, 17)⁵

Una vez más, para explicar porqué la anécdota es graciosa, nos sirve distinguir entre dos maneras de entender la pregunta ¿porqué X hace Y?: una que nos pregunta el **fin** que persigue X al hacer Y, y otra que nos pregunta porque, dado el fin que Y persigue en su acción, X es un medio adecuado para conseguirlo. La primera busca

⁵. "Someone once asked Slick Willie Sutton, the bank robber, why he robbed banks. The question might have uncovered a tale of injustice and lifelong revenge. Maybe a banker foreclosed on the old homestead, maybe a banker's daughter spurned Sutton for another.

Sutton looked a little surprised, as if he had been asked "Why does a smoker light a cigarette?"

"I rob banks because that's where the money is," he said". Traducción mía.

una respuesta que he llamado trascendental, mientras que la segunda busca una respuesta instrumental. En términos de Rawls, la primera pide un valor, mientras que la segunda pide normas. La respuesta que buscaba el reportero era trascendental, pero la respuesta que dió el ladrón fue instrumental. Si aplicamos la distinción de Rawls a epistemología, diríamos que la verdad es el valor que justifica nuestras prácticas epistémicas, mientras que una creencia está justificada si obedece las normas de dichas prácticas epistémicas.

Con esta distinción en mente es posible dar una caracterización más precisa de lo que queré decir por “infalibilismo”. El infalibilismo es la tesis de que un enunciado normativo instrumental es verdadero sólo si garantiza el cumplimiento de una regla trascendente. En otras palabras, lo instrumental debe garantizar lo trascendente. Esta caracterización hace mas precisa la idea de que, para el infalibilista, solamente son válidos aquellos medios que nos llevan de manera infalible al logro de nuestros objetivos. En contraste, el no infalibilista acepta que puede haber enunciados normativos instrumentales verdaderos pese a que no garantizan el cumplimiento de normas trascendentes.

Nótese, además, que esta noción de ‘disponible’ depende de factores muy contextuales. Como veré mas a detalle en la sesión que hablaré sobre *relativismo*, los recursos disponibles para una persona en un momento podrían ser muy distintos de los disponibles para otra o para la misma pero en otro momento. Conforme voy conociendo mas sobre una situación mas puedo prever cómo se desarrollará y actuar en consecuencia. Igualmente, los recursos disponibles para decidir cómo actuar comúnmente son muy distintos de los recursos disponibles para evaluar como se ha actuado, lo que hace que una caracterización que sea trascendentes desde el punto de vista del agente pueda ser instrumental desde el punto de vista de quién evalúa la acción del agente. Lo mismo sucede, por ejemplo, en lógica, donde la cuestión de cómo caractericé un buen argumento dependerá de si queré una caracterización que nos permita construir nuestros propios argumentos o si lo que queré es saber cómo evaluar los argumentos de otros (Morado 2000).

Quiero dejar intencionalmente impreciso qué entenderé por ‘disponible’, ‘accesible’ y bajo nuestro ‘control’ en este primer acercamiento a la distinción por dos razones. En primer lugar, porque me parece que la distinción es lo suficientemente clara por el momento y creo que a lo largo del curso se irán aclarando estas nociones en su uso; en segundo lugar, porque, pese a que en los ejemplos paradigmáticos es suficientemente claro

cuando una caracterización es instrumental y cuando es trascendente, es un debate profundo dónde dibujaré la línea entre los recursos accesibles y los inaccesibles, entre aquello que cae bajo nuestro control y qué no. En meta-ética, por ejemplo, Fernando Rudy (manuscrito) ha recientemente propuesto una teoría de las capacidades que se opone a la más tradicional propuesta volicionista y el punto central del debate es precisamente qué aspectos se encuentran bajo el control de agente en el sentido relevante para hacer evaluaciones morales: aquello que hace o puede dejar de hacer de manera voluntaria (como sostienen los volicionistas) o aquello para lo que tiene alguna capacidad (como sostiene Rudy). Quisiera permanecer neutro respecto a este tipo de debates y abierto a la posibilidad de que disponibilidad, control y acceso signifiquen diferentes cosas para casos distintos, de tal manera que para ciertos debates puede ser necesario entenderlos en un sentido muy restringido y en otros en uno más amplio. En particular, quiero trabajar con un sentido de ‘disponibilidad’, ‘control’ y ‘acceso’ que nos permita aplicarlos en casos en los que sería poco natural hablar de un **alguien** que tiene el control o acceso a dichos recursos, por ejemplo, cuando hablamos de procesos colectivos o naturales. Para ello, por supuesto, sería necesario extender también las nociones de ‘meta’ y ‘medio’.

Tradicionalmente (por ejemplo, en Ove, Edvarsson y Cantwell 2016) se presentan a las *metas* como teniendo como función típica la de

“...regular (dirigir y motivar) la acción de manera que promueva el logro de dicha meta. Bajo circunstancias normales, tener un objetivo significa que uno hará algo para lograrlo. Un agente que ha adoptado una meta realizará típicamente acciones que crea podrían acercarla a su meta y evitar las acciones que crea tendrán el efecto contrario (véase Nozick 1993 sobre principios).” (Ove, Edvarsson y Cantwell 2016, 492. Traducción mía)

En otras palabras, explicamos porqué la gente hace lo que hace asumiendo que muchas de nuestras acciones y decisiones tienen sentido y dirección, es decir, que apuntan hacia algo como meta u objetivo. Y esta presunción es lo que da sentido a nuestras prácticas normativas, es decir, precisamente porque pensamos que muchas veces hacer lo que hacer para lograr algo que tiene sentido evaluar dichas acciones respecto a su relación con la meta hacia la que están dirigidas.

Sin embargo, en este libro quiero extender la noción de meta de este, su sentido original, a uno más amplio que incluya, no sólo acciones sino también otro tipo de estados, mecanismos, herramientas, convenciones, instituciones, etc. no personales. ¿Qué quiero decir con esto? Que, en tanto hay ámbitos normativos en los cuales lo que estamos evaluando no son acciones intencionales o cosas similares, sino ámbitos en los cuales no tiene sentido siquiera hablar de un *alguien* que esté motivado y busque alcanzar dicha meta, y que, sin embargo, muestran fenómenos estructuralmente similares a aquellos en los cuales sí está involucrada una persona, vale la pena eliminar esta condición de la concepción tradicional de “meta” y trabajar con una noción mas amplia. Por supuesto, no quiero hacer revisionismo teórico o señalar que los filósofos anteriores estaban equivocados al sostener que las metas tienen una función motivacional o que están ligadas a los deseos, intenciones y otros estados mentales de una persona. Claro que no, estos elementos personales juegan un papel importante para explicar ciertos aspectos de la acción humana; pero para los objetivos de este curso, a saber, determinar la estructura de nuestras prácticas evaluativas.

Tampoco quiero meterme en cuestiones del estatus metafísico de las metas. Es un problema filosófico importante el determinar si ciertas normas o valores son constitutivos de ciertas acciones, prácticas, etc. o si para que algo sea una meta es necesario que se actúe con la intención de alcanzarla. Pongamos un ejemplo sencillo. Supongamos que alguien decide entrar a jugar golf con el objetivo de perder el juego e intencionalmente lanza la pelota lejos del hoyo en turno. ¿Cuál es la meta relevante para evaluar este acto? Al no meter la pelota al hoyo, ¿ha alcanzado el jugador su meta o no? Uno podría pensar que, en tanto el jugador no tenía la intención de ganar el juego, la meta de su acción era **no** meter la pelota. Sin embargo, también podría argumentarse que, aunque la intención era no meter la pelota, no es inadecuado decir que al no meter la pelota se ha fallado en el objetivo del *juego*. A fin de cuentas, el objetivo del juego de golf es autónomo y relativamente independiente de las intenciones individuales de quienes lo juegan.⁶ En otras palabras, para que podamos decir que lo que hace la persona es jugar golf, tenemos que aceptar como apropiado evaluar su actuar conforme a las reglas de dicho

⁶. Digo ‘relativamente’ independiente, porque en el fondo las reglas que de hecho tiene el juego también están fundadas, en última instancia, en las intenciones de quienes lo juegan en el sentido de que si los jugadores quisieran jugar el juego de otra manera, podrían convenir en otras, nuevas reglas.

juego, incluyendo las que definen su meta. La meta de meter la pelota en el hoyo es constitutiva del jugar golf, de tal manera que quien juega golf acepta de manera tácita esta meta, aunque sus intenciones vayan en dirección contraria.

La idea de que ciertas normas y metas son constitutivas de ciertas prácticas y que, por lo tanto, son apropiadas para evaluar las acciones que se realizan dentro de dichas prácticas, independientemente de las intenciones de quienes las realizan, es una idea de profundas raíces Kantianas que ha sido revivida en años recientes. En filosofía del lenguaje, por ejemplo, autores como Alston (2000), Brandom (1994), Douven (2006), Searle (1969, 2001), Sellars (1963), y Williamson (2000) han sostenido que “la aserción es análoga a un juego, en el sentido de que implica intrínsecamente ciertas reglas” (Rescorla 2009, p. 99. Mi traducción)⁷ y, en particular, que tiene entre sus reglas constitutivas el tener como meta el conocimiento (o la creencia, o la verdad, etc.). Esto significa que si afirmamos algo que no sabemos (o que no creémos, o que no es verdadero, etc.) habremos hecho algo mal, independientemente de nuestras intenciones. Alguien quien quiera decir una mentira habrá hecho algo malo si nos miente, aun si su intención era mentirnos. En meta-ética, algunos filósofos contemporáneos han defendido la posición meta-ética Kantiana de que por lo menos algunas de nuestras reglas morales genuinas son constitutivas de la acción humana libre (Korsgaard 2000), mientras que por otro lado, filósofos como Velleman (2000) y Steglich-Petersen (2006) han defendido que es constitutivo de la creencia el considerar su contenido como verdadero sólo si es verdadero.⁸

En años recientes, esta idea de que hay valores constitutivos asociados a ciertos tipos de actos, independientemente de las intenciones de quienes los realizan, ha adquirido también relevancia en ciertos

⁷. “On this view, assertion is analogous to a game, in that it intrinsically involves certain rules. Someone who exhibits no sensitivity to the rules of a game does not grasp what it is to play the game, and someone who exhibits no sensitivity to assertion’s constitutive norms does not grasp what it is to assert a proposition.” Mi traducción.

⁸. Nótese que aquí no estoy haciendo una diferencia entre lo que Steglich-Petersen (2006) llama los sentidos “teleológico” y “normativo” de las normas constitutivas relevantes. Aunque la distinción es importante, no es relevante para los objetivos de este trabajo.

debates en ética aplicada. Rae Langton y Caroline West (1999), por ejemplo, han argumentado que al consumir pornografía, independientemente de cuales sean nuestras intenciones, participamos en una práctica que tiene como meta subordinar a la mujer; mientras que Robin Zheng y Nils-Hennes Stear (manuscrito) han extendido el argumento de Langton para defender que el pintarse la cara de negro, en el contexto actual en los Estados Unidos, es participar en una práctica de opresión hacia la comunidad afro-americana, aun si nuestras intenciones son inocentes y no tenemos ninguna animadversión hacia dicha comunidad. Si bien estas discusiones tienen clara relevancia filosófica, son irrelevantes para mis propósitos, pues mi teoría se aplica tanto en los casos en los que las normas son constitutivas como en las que no, e incluso a casos en los cuales la evaluación es completamente externa a aquello que se evalúa. Por ejemplo, no tenemos que dirimir el debate de si el valor adecuado para evaluar un platillo de cocina es su sabor o su valor nutricional para reconocer que existen prácticas normativas asociadas a ambos valores y que, en cada una de ellas, tiene sentido distinguir entre metas y medios, enunciados normativos trascendentes e instrumentales, recursos disponibles para la persecución de dicho valor, etc.

D. Sesgos Cognitivos

Permítaseme entonces explicar cómo funciona este sentido extendido de ‘meta’ (y las nociones relacionadas de ‘medio’, ‘disponibilidad’, ‘acceso’, ‘control’, etc.) desarrollando un ejemplo que ya había mencionado anteriormente: los sesgos epistémicos:

Cuando recién ingresé al Instituto de Investigaciones Filosóficas, yo venía de hacer lógica, es decir, me consideraba un lógico y en mi posgrado me había dedicado principalmente a la lógica (y la filosofía de las matemáticas). En ese entonces mi concepción de la racionalidad era que ésta consistía en razonar de manera lógica, es decir, según las reglas que la lógica buscaba describir y que, en gran parte, había descrito ya con éxito. Por ejemplo, esto significa que si uno quiere determinar si una proposición de la forma *Para-toda- x -($P(x)$ -implica- $Q(x)$)* es verdadera o falsa, es necesario y suficiente eliminar la posibilidad de un contraejemplo a que sea P pero no Q . Por ejemplo, si queré saber si en un bar se respeta la regla de que sólo los mayores de edad beben alcohol (que es una regla de la forma *Para-toda- x -($P(x)$ -implica- $Q(x)$)* donde P es la propiedad de beber alcohol y Q la propiedad de ser mayor de edad), sabemos fácilmente que hay que checar a los menores de edad para ver si

están bebiendo alcohol o a los que beben alcohol para ver si son mayores de edad o no. A decir verdad, la evidencia empírica muestra que casi cualquier persona puede efectuar esta simple inferencia que manifiesta su capacidad racional. Desafortunadamente para esta concepción de la racionalidad, también es cierto que si le pedimos a la gente que realice otras inferencias *con exactamente la misma forma*, no nos dan la respuesta correcta. Por ejemplo, si poné frente a la gente cuatro tarjetas, cuyas caras visibles de las cartas muestran, las primeras dos los números 3 y 8, y las otras dos las letras R y A y les decimos que cada una de ellas tiene un número de un lado y una letra del otro. Al preguntarles qué cartas deberían dar vuelta para comprobar si es cierto que si una carta muestra un número par por un lado, entonces la cara opuesta muestra una vocal, la gente comúnmente responde que se deberían voltear la carta con el número 8 y la carta con la letra E. Sin embargo, un poco de atención nos permite dar cuenta de que la respuesta correcta (si entendé al condicional como expresando una implicación material) es que se deberían voltear la carta con el número 8 y la carta con la letra R.

Lo sorprendente de este ejemplo es que este problema tiene exactamente la misma forma lógica que el anterior y, sin embargo, mientras que el primero lo resuelve la gente de manera correcta sin muchos problemas, este segundo casi nadie lo resuelve de manera correcta aun después de mucha reflexión. Además, otro punto clave a notar era que este error no era una excepción rara que podríamos explicar apelando a un descuido o falta de pericia lógica, pues el experimento da prácticamente los mismos resultados cuando se aplica a estudiantes de lógica que cuando se aplica a otras personas. Pero lo mas importante, es que era un *error* que se reproducía de manera sistemática, lo que mostraba que era natural y, posiblemente constitutivo de nuestra forma de razonar. Pero, ¿qué debé concluir de esto? ¿Qué la gente es irracional porque no puede ni siquiera hacer una operación tan simple y básica como detectar las condiciones de verdad de un condicional? ¿O que mas bien la racionalidad no tiene que ver con reglas lógicas sino con otras cosas que mas bien sólo la psicología nos puede revelar?

Obviamente, ambas respuestas son inaceptables. Como buen estudiante de Frege, sabía muy bien que la validez lógica era independiente de cómo de hecho razonamos y que ningún experimento psicológico podría convertir una inferencia no válida en una válida, por mas que mostrara que todo mundo infiere así. Pero por otro lado, sabía también que una teoría de la racionalidad que tuviera como consecuencia que casi todos somos

irracionales tampoco podía ser una teoría correcta. Era necesario, pues entender mejor la relación correcta entre la validez lógica y la racionalidad humana.

Es aquí que la distinción entre reglas trascendentes e instrumentales nos sirve para entender la relación entre validez lógica y racionalidad. En particular, es posible concebir a las reglas lógicas formales como reglas trascendentes de la racionalidad y las que de hecho guían nuestro razonamiento como instrumentales. No es necesario entrar en detalles, pero ahora entiendo que lo que sucede en estos casos es que este tipo de “errores” lo que muestran es que nuestras capacidades cognitivas no están diseñadas para siempre producir validez formal, pues hacer esto requiere ejecutar operaciones muy complejas (por ejemplo, detectar la forma lógica de la información que se nos presenta) que, además, nos tomaría demasiado tiempo. En la mayor parte de los casos, y en los más relevantes para nuestra supervivencia, no tenemos el tiempo para realizar dichas operaciones y debemos distribuir nuestros recursos cognitivos (memoria, atención, etc.) entre muchas tareas cognitivas, de tal manera que comúnmente no es posible dedicar demasiados recursos a una sola inferencia. Por eso, nuestros mecanismos están diseñados para procesar información de manera eficiente, es decir, rápida y usando recursos cognitivos mínimos a costo de sacrificar cierta efectividad. En otras palabras, aun cuando en principio podríamos actuar y tomar decisiones siempre de manera lógicamente correcta, los recursos necesarios para hacerlo no nos están disponibles y por eso nos sirve de mucho contar con mecanismos cognitivos que nos permiten realizar inferencias que, no siempre son lógicamente válidas, pero son lo suficientemente certeras en suficientes casos importantes, a partir de recursos que sí tenemos accesibles. A esto me refiero cuando digo que las reglas de la lógica formal son trascendentes, mientras que nuestros mecanismos cognitivos siguen más bien reglas heurísticas que son instrumentales, pues están diseñadas para optimizar nuestros recursos cognitivos reales. Una vez más, en definir que entendí por ‘disponible’ es importante tomar en cuenta el aspecto económico que mencionamos en la sección anterior. Cuando nuestros recursos son limitados (y nuestros recursos cognitivos son, de hecho, limitados) hay ventajas importantes (en este caso, evolutivas) en contar con procesos y mecanismos menos demandantes. Esto significa que, aun cuando en un sentido absoluto tenemos disponibles los recursos necesarios para realizar, por ejemplo, inferencias lógicamente impecables (y por eso es que es posible hacer, por ejemplo, pruebas lógicas

absolutamente válidas), no siempre es posible darnos ese lujo y en este sentido dichos recursos no deben contarse como realmente disponibles en el sentido relevante.

De esta manera es posible dar cuenta de lo que sucede en el caos cognitivo de una manera que evite ambos cuernos del dilema. Por un lado, reconocí la validez lógica como valor con fuerza normativa a la hora de evaluar nuestras inferencias, pero la restringimos a casos en los que hay pocas restricciones de acceso y control, es decir, cuando está justificado dedicarle los muchos recursos de tiempo, control y atención que se necesitan para ejecutarla. Por el otro, evitamos también la conclusión apresurada de que somos constitutivamente irracionales al desarrollar una concepción de la racionalidad acotada, donde por lo menos algunos de nuestros errores lógicos no se deben a fallas de ejecución o al descuido, sino que son plenamente racionales.

En este caso se ve claro cómo nos sirve extender el lenguaje de ‘disponibilidad’, ‘acceso’, ‘metas’ y ‘medios’ a casos donde literalmente no hay una persona de la que decimos adopta un medio o un fin, de la que decimos tiene disponibles ciertos recursos limitados que debe administrar. No quiero comprometerme a que esta manera de usar estos términos sea metafórica o no, siempre y cuando se reconozcan las fuertes semejanzas entre el caso personal y el caso no-personal, de tal manera que las conclusiones que alcance de ahora en adelante puedan aplicarse tanto a unos casos como a los otros.

E. Falibilidad mas allá de la epistemología

Hoy en día, en filosofía, cuando se habla de falibilismo, se piensa automáticamente en la tesis epistemológica según la cual, “ninguna creencia (teoría, concepción, tesis, etc.) puede nunca tener un apoyo o justificación racional concluyente [es decir,] nunca podré excluir por completo el poder dudar de su verdad.” (Hetherington 2017. Mi traducción) Esta tesis a veces suele formularse apelando a la noción misma de falibilidad así: “Toda creencia está justificada de manera falible” es decir, si una creencia está justificada, esta justificación es falible, es decir, puede fallar. Así formulado el falibilismo, queda claro que la falibilidad no es una propiedad de la creencia en sí misma, sino de su justificación o apoyo racional. Lo contrario de lo falible, entonces, es lo infalible. Una creencia estaría justificada de manera infalible si su justificación por sí misma garantizara por completo su verdad. Durante mucho tiempo se creía que, por lo menos para cierto tipo de creencias podría lograrse una

justificación de este tipo. Para Descartes, por ejemplo, nuestras ideas claras y distintas estaban justificadas de manera infalible, es decir, su propia claridad y distinción garantizaban por completo su verdad.

Espero sea claro, a estas alturas del curso, que la falibilidad de la justificación de nuestras creencias es sólo un caso de un fenómeno mas general que puede darse en otros ámbitos además del epistemológico. Que algo sea falible no significa mas que puede fallar y que algo sea infalible tampoco significa entonces más que garantiza su objetivo de manera concluyente. Así pues cualquier actividad que tenga un fin u objetivo, puede ser falible – si es posible que no alcance su objetivo – o infalible – si lo garantiza.

Pongamos un ejemplo de interés filosófico: la posesión de conceptos. Es posible entender el debate sobre si es necesario para poseer un concepto dado el ser capaz de entenderlo por completo o no como un debate sobre la falibilidad de la posesión de conceptos. En otras palabras, si la posesión de conceptos fuera infalible, quien posee un concepto deber ser siempre capaz de evaluar de manera acertada, dada suficiente información sobre un objeto cualquiera, si cae o no dentro de su extensión, y ser capaz de efectuar de manera competente las inferencias necesarias que definen las relaciones lógicas del concepto – por ejemplo, inferir de que algo es un buey a que no es una vaca hembra o de que algo es un lobato a que es un lobo joven. En contraste, si la posesión de un concepto fuera falible, entonces sería posible poseer un concepto y fallar al aplicarlo, ya sea en inferencia o clasificación. (Earl 2017)

Sin embargo, alguien podría pensar que no es sorprendente que tenga sentido hablar de falibilidad aquí, pues el fenómeno de la posesión de conceptos sigue siendo un fenómeno epistemológico (Earl 2017). Sin embargo, creo que debe ser importante dejar claro que no siempre que hablamos de falibilidad lo hacemos en un plano epistemológico. Para ello, sería útil tratar de encontrar un ejemplo no epistemológico. Para lograr esto, es mas fácil empezar fuera de la filosofía. Consideré el siguiente caso de la vida real:

Hace unos días, un camión de gran peso embistió un mercado navideño en Berlín, asesinando a 12 personas e hiriendo a casi 50. Meses antes, otro camión de gran peso fue usado en la ciudad de Niza para un ataque similar que resultó en 86 personas fallecidas. En condición de anonimidad, unos días después del ataque en Berlín, un alto funcionario de fuerzas antiterroristas declaró a la *Agence France-Presse* que no existían medidas infalibles contra este tipo de ataques, aunque era posible tomar precauciones como prohibir la circulación de

automóviles y colocar bloques de hormigón alrededor de las calles donde se esperan se junten multitudes. En otras palabras, aunque las medidas podrían evitar ciertos ataques, no pueden garantizar que este tipo de ataques dejen de suceder. "Desgraciadamente, habrá que acostumbrarse a vivir bajo amenaza" declaró a la *AFP* Ludovic Guérineau, ex miembro de los servicios secretos franceses y ahora director de operaciones de la firma de seguridad privada Anticip.

En este caso, es posible hablar de medidas falibles o infalibles, no porque haya una dimensión epistemológica, sino porque hay un objetivo y este puede estar garantizado por la medida o no. Por supuesto, alguien podría re-formular el ejemplo de manera epistemológica, por ejemplo, diciendo que lo que hace a las medidas sugeridas a la *AFP* falibles es que, dado que *no es posible saber* cuando o donde habrá un intento de ataque con camiones, no es posible evitar por completo, o que el tomar dichas medidas *no nos permite saber con certeza* que no habrá ataques con camión; sin embargo, espero sea claro que el lenguaje epitémico en esta formulación del fenómeno es completamente artificiosa y no añade nada a decir simplemente que las medidas serían adecuadas aun si no eliminaran por completo la amenaza de un ataque de ese tipo. Lo que hace a una medida, técnica, proceso, etc. falible o no no es el de si, al aplicarlo, sabré o no que logrará su objetivo, sino que si dicho objetivo está o no garantizado. Si tengo razón, entonces, todo medio puede ser falible o infalible dependiendo de si garantiza o no su fin. Así es como voy a entender la falibilidad de aquí en adelante. [Agradezco a Ricardo Mena el presionarme en la importancia de este punto.]

En mi trabajo de (Barceló 2015) presenté otro ejemplo de relación filosófica importante que es útil pensé en términos de falibilidad en vez de infalibilidad: la relación entre una verdad y su base metafísica, es decir, aquello que la hace verdadera. Mi ejemplo ahí era la verdad de que mis sobrinos son traviesos, lo cual es verdadero porque mis sobrinos, Aimee y Balam, de hecho son traviesos. Sin embargo, la relación es falible – o derrotable que es otra manera de decir lo mismo, como veré mas adelante – porque el que mis sobrinos Aimee y Balam sean traviesos no garantiza por completo que mis sobrinos sean traviesos, ya que sigue existiendo la posibilidad de que tuviera otro sobrino que no fuera travieso. En ese caso, seguiría siendo el caso que mis sobrinos Aimee y Balam sean traviesos y, sin embargo, sería falso que mis sobrinos son traviesos.

Podría parecer muy extraño el que use el ejemplo de la relación entre hechos y verdades aquí dado que he

dicho desde el principio que lo que busco en este curso es presentar una teoría sobre la estructura de nuestras prácticas normativas y no hay nada normativo sobre esta relación. Lo acepto. Sin embargo, como quedará cada vez mas claro, mi meta es aun más ambiciosa y abarca cualquier tipo de relación que se da entre lo determinado y lo indeterminado y que, por lo tanto, es falible.⁹ Esto significa que se aplica también a relaciones como la de causalidad, inferencia, consecuencia, fundamentalidad metafísica, referencia, etc. Los casos normativos son paradigmáticos, pero no son los únicos.

Si bien la infalibilidad es una virtud extrema de las herramientas, reglas, convenciones, métodos, procesos, mecanismos, esquemas de comportamiento, medios, etc.¹⁰ esto no significa que las herramientas (métodos, medios, etc.) falibles sean inútiles.¹¹ Es posible que una herramienta sea falible y aun así sea útil y su existencia valga la pena. Aun cuando las medidas sugeridas a la AFP arriba no garanticen detener los ataques a civiles por medio de camiones, esto no significa que no deban de adoptarse. Pensar de otra manera es caer en lo que se llama la falacia de confundir lo bueno con lo perfecto. Una medida puede ser buena, aunque no sea perfecta, es

⁹. Esto no significa que la implicación se de en ambos sentidos. Es posible que haya falibilidad aun en casos en los cuales está completamente determinado qué debé hacer para lograr nuestros objetivos con certeza, pero por razones de escasez de recursos no sea una buena idea hacerlo.

¹⁰. En general, para ser mas conciso a lo largo del curso hablo simplemente de “medios” para referirme a cualquier miembro de este heterogéneo grupo. En este sentido, un medio es cualquier cosa cuyo comportamiento tratamos de explicar teleológicamente, es decir, apelando a algún fin que de sentido a su comportamiento.

¹¹. Cuando nos preguntamos por la utilidad de una herramientas, es importante distinguir entre tres preguntas relacionadas pero sustancialmente distintas:

1. ¿Por qué es útil que **exista** la herramienta – es decir, no sólo por qué fue inventada, sino por qué ha seguido siendo tan exitosa como para haber sobrevivido?
2. ¿Por qué es útil **poseer** una herramienta?
3. ¿Por qué es útil **usar** una herramienta en una circunstancia dada?

Continuando con el ejemplo del martillo, no es lo mismo preguntarse porqué existen los martillos, que preguntarse porqué tenemos uno en nuestra casa o porqué lo usamos para clavar algo en una ocasión dada.

decir, infalible. Pero uno cae en la misma falacia si piensa que todo método falible vale la pena. Cuando es racional o no usar un método falible es una cuestión fundamental para entender la falibilidad y de ello hablaré en el siguiente capítulo. En él veré cómo tomarse en serio el fenómeno de la suerte requiere una transformación sustancial de nuestra concepción del error, en general, y del error lingüístico en particular.

III. Riesgo y Suerte

A mediados del siglo pasado, el re-descubrimiento de dos fenómenos importantes en el campo de la epistemología y la ética obligaron a los filósofos interesados en estos campos normativos a tomarse en serio el tema de la suerte. El primero de ellos es lo que ahora conocí como los “casos Gettier” en honor al filósofo que los presentó, de manera muy elegante y sugerente, en un breve artículo de título “¿Es el conocimiento creencia verdadera justificada?” (Gettier 1963), el segundo es el fenómeno de la suerte moral, introducido a la meta-ética por Thomas Nagel (1979) y Bernard Williams (1981). Aún cuando, en realidad, ambos fenómenos no son sino casos particulares de un fenómeno más general, nos acercaré al fenómeno primeramente a través del argumento de Gettier en la siguiente sección.

A. ¿Cuál es la moral del argumento de Gettier?

Como habíamos señalado en sesiones anteriores, una de las funciones centrales de los discursos normativos es guiarnos en nuestra toma de decisiones y deliberación, tanto privada como pública. Nos involucramos en prácticas normativas precisamente en aquellos entornos en que hay algo en juego, aquellos en los cuales no solamente hay algo que ganar, sino también algo que perder. Esto hace que los discursos normativos cobren un papel importante en circunstancias en las que hay un riesgo, y lo que esperamos de un medio adecuado es una **reducción sustancial de dicho riesgo**. El actuar responsable es el hacer lo suficiente, dadas nuestras capacidades y recursos, para reducir riesgos para nosotros mismos y para aquellos que puedan ser afectados por nuestros actos. En otras palabras, nos preguntamos qué hacer precisamente en aquellas circunstancias en las que **no da lo mismo** qué hagamos, sino que cómo actué puede tener consecuencias de gravedad tal que justifiquen le dediqué cuidado a nuestras decisiones. Existe la normatividad epistémica, por ejemplo, porque no nos es indiferente qué pensé, es decir, cuales sean nuestras creencias, nuestras inferencias, nuestras dudas, etc. Pensar de manera incoherente, no aprovechar nuestra información para tener una imagen adecuada de nuestra realidad, etc. son cosas que pueden tener consecuencias negativas, no sólo para nuestra supervivencia como individuos, sino también como especies, por ejemplo. Es por eso que nos preguntamos qué debé de hacer, como debé de

actuar, para ser epistémicamente responsables, con nosotros mismos y con otros. Eso es lo que da sentido a la epistemología. De la misma manera, el lenguaje también tiene una dimensión normativa porque es importante seguir ciertas reglas para poder aprovechar la existencia del lenguaje para alcanzar nuestros fines colectivos e individuales. Si no importara en absoluto como usáramos las palabras, no habría normatividad en el lenguaje. Pero claro que importa. Si no pudiéramos esperar que los otros usasen las palabras de acuerdo a su significado, no podríamos usarlas para comunicarnos, por ejemplo.

La normatividad tiene sentido, por lo tanto, en los espacios en los que no es posible dejar las cosas meramente a la suerte, sino que buscamos algún tipo de certeza (en el caso del infalibilismo), o por lo menos confianza (en el caso no infalibilista), entendiendo ambos “certeza” y “confianza” en un sentido no-epistemológico. En este sentido, un método, medio, herramienta, etc. es adecuado para perseguir ciertos fines, si, a partir de recursos que están de hecho disponibles (es decir, que no son muy costosos, de los que tenemos acceso y están bajo nuestro control) reduce el riesgo de falla.

Ahora bien, una vez que reconocí que los medios pueden ser falibles, es decir, que un medio puede ser adecuado aun cuando no garantice al cien por ciento su fin, uno podría pensar que tenemos cuatro opciones:

1. Usamos un medio adecuado y obtené nuestro fin.
2. No usamos un medio adecuado y no obtené nuestro fin.
3. Usamos un medio adecuado y, sin embargo, no obtené nuestro fin.
4. No usamos un medio adecuado y, aun así, obtené nuestro fin.

Para el infabilista, la opción (3) no es genuina. Un medio es adecuado sólo si nos lleva a obtener nuestro fin. Pero el infabilista y el fabilista comúnmente no dicen nada sobre (4). En esta sección del curso veré que esta división de cuatro tipos de situaciones no es suficientemente fina para dar cuenta de nuestras verdaderas prácticas normativas. Veré que no todos los casos de tipo (1), por ejemplo, son normativamente equivalentes, por ejemplo, veré que no siempre que se da (1) es posible hablar de un logro. Esta es la conclusión del famoso argumento de Gettier.

El argumento de Gettier se basa en dos premisas sencillas: Primero, que la justificación es falible, es decir, que es posible que una persona esté justificada en creer una proposición que es de hecho falsa (con los ajustes

adecuados para las proposiciones necesarias);¹² en segundo lugar, que la deducción (competente) extiende la justificación, es decir, para cualquier proposición P , si S está justificado en creer P , y P implica lógicamente Q , y S hábilmente deduce Q de P y acepta Q como resultado de esta deducción, entonces el que S crea que Q está justificado. Presumiblemente, ambas premisas se aplican tanto a la justificación *a priori* como a la *aposteriori*.

Para aquellos que puedan no estar familiarizados con el argumento de Gettier, permítanme ofrecer una reconstrucción muy general y abstracta (o, al menos, una reconstrucción de su caso disyuntivo II, el caso conjuntivo I es completamente análogo y dual):

- 1 Hipótesis a reducir: El conocimiento es creencia verdadera justificada
- 2 Premisa 1: La justificación es falible. Es posible que una persona esté justificada en creer una proposición que es de hecho falsa. Sea S tal persona y p la proposición falsa que S cree justificadamente.
- 3 Sea q otra proposición que S no sabe, aunque sea verdadera.
- 4 p implica lógicamente $(p \vee q)$.
- 5 Supongamos que S ve la implicación de p a $(p \vee q)$, y acepta $(p \vee q)$ en base a (p) .
- 6 Premisa 2: La justificación se transmite a través de, al menos deducciones lógicas básicas como la que va de p a $(p \vee q)$.
- 7 De 1, 5 y 6, S está justificado a creer $(p \vee q)$.
- 8 q implica lógicamente $s(p \vee q)$.
- 9 De 3 y 8, $(p \vee q)$ es verdadero.
- 10 De 7 y 9, S está justificado a creer la proposición verdadera $(p \vee q)$.
- 11 Sin embargo, $(p \vee q)$ es verdadero porque q es verdadero (de 3, 8 y 9), y S no sabe que q es verdadero (de 3).
- 12 De 11, S no sabe que $(p \vee q)$.
- 13 Por lo tanto, S está justificado en creer una proposición verdadera $-(p \vee q)-$ que no conoce, lo que contradice nuestra hipótesis inicial (1).

¹². En el sentido relevante de "justificado" en el cual el que un agente S esté justificado en creer que P es una condición necesaria de que S sepa que P

El argumento es perfectamente sólido, por lo que la única manera de resistirse a su conclusión sería atacar una de sus premisas. La premisa 2 parece razonable sin duda, pero ¿qué razones debé tener para aceptar la primera premisa, es decir, la falibilidad de la justificación? Gettier no lo dice explícitamente, pero claramente parece una presuposición de la definición misma de conocimiento que se discute. Para que el conocimiento sea una verdadera creencia justificada, cada elemento de la definición debe hacer algún trabajo. Si la justificación no fuera falible, implicaría la verdad, y ésta sería descartable de la definición. Además, puesto que la justificación presupone la creencia, justificarse sería una condición necesaria y suficiente para el conocimiento. En otras palabras, la justificación no sería más que una palabra rara para hablar del conocimiento y como tal no podría definirla. En consecuencia, la justificación no puede implicar verdad, es decir, debe ser falible.

Algunos filósofos han considerado que la moraleja de los argumentos de Gettier es que hay un ingrediente o condición extra que falta ser añadido a la justificación y la verdad para obtener conocimiento. Sin embargo, uno debe tener cuidado de no agregar ninguna condición adicional que también se transfiera a través de deducciones lógicas inmediatas y sea falible, pues el argumento de Gettier también se aplicaría a tal condición adicional. Supongamos que alguien sostuviera la hipótesis de que el conocimiento no es una creencia verdadera y justificada, sino una creencia verdadera justificada más alguna otra condición. Llamé a esta condición adicional C . Ahora, dado que C es falible, es posible que alguien S tenga una creencia falsa p que satisfaga C . Toma de nuevo otra proposición verdadera q que S no sepa. Si S no está justificado ni en creer p ni en creer q , toma entonces una tercera proposición r que S crea de manera justificada, pero no sepa. Una vez más, si S hace la inferencia de p a $(p \vee q)$ o $((p \vee q) \vee r)$, tendría una creencia verdadera justificada que también satisface la condición C de la que, sin embargo, no tiene conocimiento. Por lo tanto, ninguna tercera condición de este tipo puede ayudar a construir el puente entre la creencia verdadera justificada y el conocimiento. Esto significa que si hubiera una nueva condición que añadir esta tendría que ser una que no se transmitiera por deducciones básicas competentes – lo cual sería muy raro, aunque hablaré de ello en el último capítulo de este libro – o que fuera infalible. Sin embargo, ya he visto que sería difícil encontrar una condición instrumental – es decir, una que hiciera más que reformular la noción de conocimiento, sino que nos ayudara a guiar nuestro comportamiento – que fuera infalible. Nos queda la tesis negativa de que la creencia verdadera y justificada no es conocimiento, puede ser una

condición necesaria, pero no suficiente. Pero la probabilidad de encontrar una condición adicional para completar el análisis es muy baja, por lo que el conocimiento no puede ser analizado en términos de justificación.

Nótese que aunque el argumento está tradicionalmente formulado en términos de conocimiento proposicional, es decir, conocimiento **qué**, es también válido cuando se aplica al conocimiento práctico, es decir, al conocimiento **cómo**, pues éste también es falible. En este caso, la conclusión relevante sería que si algo es el tipo de cosa que alguien podría saber cómo hacer, entonces es posible hacerlo, aun para quienes saben como hacerlo, no como manifestación de que saben hacerlo, sino por mera suerte.

Premisas:

1. Saber cómo es falible, es decir, un agente puede saber cómo hacer X y, sin embargo, fallar al intentar hacer X .
2. Si hacer X es una forma de hacer Y , entonces saber cómo hacer X implica saber cómo hacer Y .

Argumento:

3. Sea a un agente que sepa hacer X and sin embargo falle al intentar hacerlo.
4. Sea Y algo que a hace, pero sin saber cómo hacerlo.
5. Hacer X y hacer Y son maneras de hacer (X o Y).
6. a sabe como hacer (X o Y) (por 2, 3 y 5), hace (X o Y) (por 4 y 5) y, sin embargo, su hacer (X o Y) no manifestó su capacidad de hacer (X o Y).

Por lo tanto, a hace (X or Y) por pura suerte.

La idea fundamental detrás del argumento es que si hay (por lo menos) dos maneras distintas de hacer algo, basta que uno sepa hacerla de *alguna* manera para que sepa hacerla; esto significa que es posible que uno, que sabe hacerla de una manera, intente hacerla de dicha manera y al fallar, por suerte, terminé haciéndola de la otra. De esta manera, habrá hecho algo que sabe hacer, aunque por suerte. Esto demuestra que algunas cosas que se puede saber cómo hacer, también se pueden hacer (incluso por los que saben cómo hacerlo) por pura suerte.¹³

Aun más, creo que el argumento de Gettier es lo suficientemente amplio como para aplicarse no sólo al conocimiento sino a nociones normativas similares, es decir, aquellas en las que existe una meta (en este caso, la

¹³. Agradezco a Miguel Ángel Sebastián el haberme corregido sobre la formulación correcta de la conclusión de este argumento.

verdad) y un medio para alcanzar dicha meta que puede ser evaluada positivamente pese a no ser infalible. En otras palabras, el argumento de Gettier es pertinente siempre que hay riesgos involucrados, como lo desarrollaré en la siguiente sección.

B. Riesgo y Normatividad

"El futuro tiene muchos nombres.

Para los débiles es lo inalcanzable.

Para los temerosos, lo desconocido.

Para los valientes es la oportunidad".

Victor Hugo

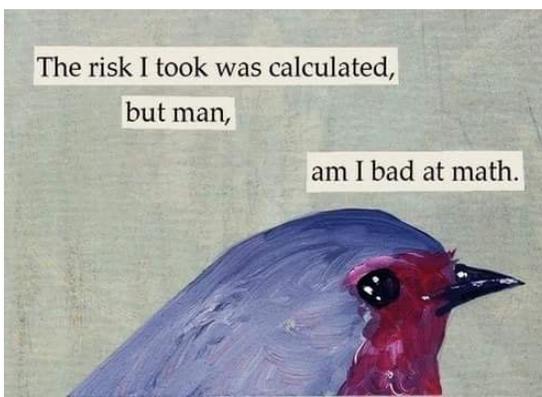
Cada vez que se toma un riesgo, existen por lo menos dos maneras de evaluar el acto: uno, juzgar si estuvo justificado asumir el riesgo (dado el estado de conocimiento del sujeto, la información disponible, las probabilidades, las ganancias posibles de los posibles resultados, etc.) y dos, simplemente juzgar si el resultado fue el deseado. Arriesgar todo nuestro patrimonio en una carrera de caballos de cuyos participantes no sabemos nada es insensato, aun cuando gané la puesta y la riqueza que obtengamos de la misma nos traiga enormes beneficios. Inversamente, hay un sentido en el que, aún cuando estamos justificados en tomar un riesgo, *post facto* sentimos que estuvo mal tomarlo si el resultado que obtuvimos no fue el deseado.¹⁴ Distinguir entre estos dos tipos de juicios es uno de los puntos centrales de Bernard Williams en su seminal artículo "Suerte Moral" (1981). Cada vez que tomamos el volante de un auto, estamos corriendo el riesgo de atropellar a alguien. Si somos cuidadosos al manejar, reducimos este riesgo, pero no siempre lo eliminamos. Es decir, es posible manejar con cuidado y aun así atropellar a alguien. Cuando esto sucede, tenemos razón en pensar que hicimos algo horrible, aun cuando no podamos ser acusados de negligencia al no encontrarse ningún descuido o irresponsabilidad de nuestra parte. La primera intuición (la de que hicimos algo horrible, a saber, atropellar a alguien) es capturada por el primer criterio normativo, mientras que el segundo criterio busca capturar la intuición detrás del segundo

¹⁴. Desde el punto de vista psicológico, la evidencia parece indicar que ambos tipos de riesgos son evaluados por mecanismos cognitivos distintos, por lo que presumiblemente tienen historias evolutivas distintas (Cushman 2015).

juicio (de que somos inocentes, pues actuamos de manera cuidadosa y responsable). Cuando un acto es impugnable según el primer criterio, pero no según el segundo, tenemos un desafortunado caso de *mala suerte*.

Como ya he sugerido, esta distinción es muy importante para la discusión ética contemporánea, pues se encuentra al centro de la discusión alrededor de la suerte moral (Nagel 1979, Williams 1981). Sin embargo, también es importante para la epistemología y la semántica filosófica. Para entender su importancia en epistemología, basta reconocer que toda creencia involucra un riesgo. Si he aprendido algo de veinticinco siglos de escepticismo filosófico en occidente es que no es posible garantizar por completo la verdad de ninguna o prácticamente ninguna de nuestras creencias. En otras palabras, aunque sea de manera implícita, cada creencia nuestra involucra el riesgo de que lo que creamos no sea verdadero. Si nuestra creencia es justificada, he satisfecho los requisitos que impone el primer criterio de normatividad (he tratado de reducir el riesgo de falsedad, cumplido nuestras obligaciones epistémicas, etc.). Si nuestra creencia es verdadera, he satisfecho el segundo. La diferencia entre verdad y justificación, por lo tanto, no es sino la versión epistemológica de la misma distinción general entre ambos tipos de juicios.

Poner la distinción en términos epistémicos nos sirve además para recordarnos que existe un tercer criterio normativo involucrado en nuestra evaluación de actos arriesgados. Si bien es cierto que hay algo malo en tomar riesgos injustificados, o que algo malo sucede si no obtengo el resultado deseado al correr un riesgo, como Gettier (1963) ya bien señaló, muchas veces no es suficiente estar justificado en asumir un riesgo *y además* obtener el resultado deseado, sino que también es deseable que la satisfacción de ambos criterios esté adecuadamente conectada. En otras palabras, para satisfacer este tercer criterio normativo, no es suficiente que uno haya tenido el cuidado suficiente para reducir los riesgos a un grado razonable, o que se haya obtenido el resultado que se esperaba, sino que es necesario también que dicho resultado se haya obtenido precisamente porque se tomaron dichos cuidados. El contraste aquí es con actos arriesgados (posiblemente cuidadosos) en los cuales el resultado



esperado se obtiene por *buena suerte*.

La distinción también es importante en semántica, pues todo lo dicho hasta ahora sobre la creencia se aplica *mutatis mutandi* también a la afirmación. Afirmar siempre involucra asumir un riesgo: el riesgo de que lo que digamos resulte ser falso. Esto es lo que significa decir que la verdad es la norma de la aserción. Sin embargo, en tanto que hay un riesgo involucrado, las tres fuentes de normatividad anteriormente mencionadas vuelven a aparecer: una afirmación puede estar justificada, ser verdadera, o ser verdadera gracias a que está justificada. Idealmente, una afirmación debe satisfacer los tres criterios; pero si no es así, uno debe ser cuidadoso distinguir las diferentes razones por las cuales la afirmación no es óptima. (En 2012, y en sesiones posteriores de este mismo curso, argumento que ignorar esta distinción es un tipo de error detrás del relativismo de John MacFarlane (2014), por lo menos, respecto a modalidades epistémicas). Es por eso que retractarse de una afirmación no es lo mismo que afirmar que lo dicho era falso. Uno puede también retractarse porque lo dicho no estaba justificado, por ejemplo. No sé si existen casos en los que uno deba retractarse de una afirmación gettierizada, es decir, verdadera, justificada, pero donde la verdad no esté adecuadamente vinculada a la justificación, pero supongo que sí.

Como mencioné al final de la sección pasada, la distinción se aplica también fuera de la epistemología en cualquier situación en la que hay un riesgo – calculable o no. Permítanme ilustrar esto con un ejemplo. Supongamos que efectivamente en Berlín se decide colocar bloques de hormigón alrededor de las calles donde se esperan se junten multitudes para impedir otro ataque con camiones. Como bien señalaba el oficial anónimo de las fuerzas anti-terroristas, esto es una medida razonable aun cuando no nos pueda garantizar al cien por ciento que no sucederá otro ataque de este tipo. Después de todo, sigue siendo posible que los terroristas decidan atacar una zona sin bloques de hormigón. Si esto sucede, los bloques de hormigón no servirán de nada. En este sentido, aun cuando se coloquen los bloques de hormigón, no se ha eliminado el riesgo de ataque en camión. Aun así, esto no significa que es seguro que habrá un ataque y que este será exitoso en la zona donde no hay bloques de hormigón. Si había planes para atacar una zona donde se pusieron bloques y estos sirvieron para disuadir a los atacantes de intentarlo, podré decir que fue un logro de las fuerzas anti-terroristas el evitar dicho posible ataque. Pero si el ataque sucede en otra area, su éxito o fracaso ya no tendrá nada que ver con los bloques sino con

factores externos. Quedará a la suerte. Puede ser que los atacantes sean tan incompetentes que el ataque falle. En ese caso podríamos decir que las fuerzas anti-terroristas tuvieron buena suerte. Pero si mas bien lo logran, no podré acusar a dichas fuerzas de negligentes, sino que lo más adecuado sería decir que se tuvo mala suerte.

C. ¿Qué es lograr algo?

Los temas de la suerte y los accidentes han alcanzado una nueva importancia en años recientes debido al proyecto de ciertos filósofos de definir la agenda epistemológica en estos términos. La idea básica es sencilla: No se puede SABER por accidente; lo que sí se puede es tener una creencia que sea verdadera por accidente. Por ejemplo, si un estudiante copia mal en un examen de matemáticas, pero por suerte le atina el resultado correcto, es posible decir que tiene una creencia matemática verdadera que no es conocimiento.

Aunque los diccionarios los presenten como sinónimos, intuitivamente para lograr algo, no basta conseguirlo, obtenerlo o alcanzarlo. En otras palabras, una vez que nos proponé algo como objetivo o emprendé una tarea, no basta tener éxito y conseguir eso que nos proponé para decir que lo logramos. Es necesario también que el logro no haya sido accidental o por suerte, sino el resultado directo de ejercer alguna capacidad o hacer un esfuerzo dirigido a alcanzar dicho fin. Por ejemplo, si alguien se propone comprar un coche nuevo, se pone a trabajar ahorrar, etc. hasta que después de muchos años de trabajo se gana el coche en una lotería, no diríamos que por fin *logro* obtener el coche que quería. Podríamos decir que por fin *obtuvo* lo que quería, pero no diríamos que hubo ningún logro de su parte. Tuvo suerte, buena suerte.

Hasta el momento, esta concepción de lo que es un logro no es controversial. La disputa filosófica empieza a partir de aquí. En general, dos preguntas permanecen abiertas: la cuestión *descriptiva* de explicar en qué consiste exactamente el que lograr un objetivo se deba al ejercicio de alguna capacidad o sea el resultado directo de hacer un esfuerzo para alcanzarlo, es decir, explicar exactamente qué tipo de relación debe haber entre el esfuerzo o ejercicio de una capacidad y la obtención del objetivo para decir que alcanzar dicho objetivo fue un logro y no un afortunado accidente, y la cuestión *fundacional* de porqué valoramos más un logro (debido al esfuerzo propio o al ejercicio de alguna capacidad) que el obtener lo que deseamos de manera accidental o por

buena suerte, mientras que castigamos igual los casos en los que el mal es resultado de mala suerte que cuando no lo es (Martin & Cushman 2015). Tratar de responder esta segunda pregunta fundacional ha sido uno de los objetivos centrales del trabajo de filósofos como John Grecco, Peter Graham, Ernest Sosa, etc. La cuestión es especialmente complicada porque al tratar de definir un logro como, por ejemplo, aquel que se debe al ejercicio de una capacidad virtuosa, el problema pasa a determinar (de una manera no circular) qué es una capacidad virtuosa y cuando se ejerce. Una vieja propuesta de solución ha sido tratar de definir suerte y/o accidente de manera independiente y luego definir el logro como la realización no accidental de una tarea. El esfuerzo más conocido en esta dirección tal vez sea el de definir conocimiento en términos de seguridad (Sosa 2007). Bajo esta perspectiva general, un resultado es accidental si no es seguro que hubiera pasado en circunstancias similares (o mas bien, similares en los aspectos relevantes). Un resultado es seguro, en este sentido, si depende condicionalmente del esfuerzo que se hizo para obtenerse, es decir, si en los mundos posibles adecuados (adecuados para el ejercicio de la aptitud relevante) en que se hace el esfuerzo, se da el resultado. Por ejemplo, un experto sexador de pollos no identifica por accidente el sexo de un pollo, ya que sería capaz de identificar el sexo del mismo pollo (u otro suficientemente parecido) en casi cualquier circunstancia similar. En otras palabras, esperamos (de manera derrotable) que sea raro que quién esté bien preparado fallé, que si tenemos suerte, ésta sea endeble, en vez de robusta, es decir, que pequeños cambios cambien el resultado positivo o negativo.

Otras propuesta más reciente de resolver el problema descriptivo ha sido tratar de definir las habilidades como un tipo de disposición, y luego definir el logro en términos del ejercicio de una habilidad. En esta propuesta, una persona tiene la habilidad de realizar una tarea si tiene la disposición a realizar exitosamente dicha tarea en las circunstancias relevantes con un grado sustancial de confianza (Greco 2010). Bajo esta concepción de lo que es una habilidad, una persona puede no lograr realizar una tarea por una de tres razones: porque no tiene la habilidad de realizar la tarea, porque las condiciones necesarias para el ejercicio de la habilidad no se dan o porque -- aunque el sujeto tenga la habilidad y las circunstancias sean las adecuadas -- dicha habilidad sigue siendo falible. Finalmente, también se ha tratado definir la posesión de una capacidad (o por lo menos, de nuestras capacidades básicas) en términos evolutivos: tenemos la habilidad de lograr ciertas tareas porque uno o varios de nuestros órganos (u otros rasgos fenotípicos) fueron seleccionados con dicha

función. En este sentido, por ejemplo, tenemos la habilidad de recordar hechos pasados (en las circunstancias relevantes, y de manera falible) porque parte de nuestro aparato cognitivo fue seleccionado para realizar esta función. Sin embargo, aun cuando este tipo de estrategias puedan servir para explicar algunas de nuestras capacidades – las mas “naturales” por decirlo de alguna manera –, dudo que sea viable encontrar una explicación adaptativa para todas ellas.

Por mi parte, yo prefiero modelar la noción de justificación en términos de “estar preparado” en vez de “tener la capacidad” o “ejercer una habilidad”, porque esta noción claramente revela aspectos importantes de la justificación, algunos de los cuales filósofos de la virtud como Sosa y Grecco habían ya señalado. En particular, porque revela que, detrás de la noción absoluta de justificación – según la cual, o bien se está justificado/se está preparado/se tiene la capacidad o no se tiene – hay otra noción relativa – según la cual, uno está preparado para una situación o no, relacionadas de tal manera que uno está preparado (en el sentido absoluto) cuando se está preparado (en el sentido relativo) para no sólo las situaciones que más se esperan, sino también para enfrentar por lo menos algunas eventualidades **previsibles**. ¿Cuáles y cuantas? Las suficientes, y aquí, por supuesto, la noción de suficiente convierte a la noción de justificación en un aspecto gradual: uno puede estar más o menos justificado si actuó preparado para más o menos situaciones que se pudieran presentar.

Ha habido intentos de modelar la suerte de manera modal o probabilista, pero han fallado, pues aunque son nociones relacionadas, la suerte es determinante sólo en casos imprevisibles y lo previsible en este sentido no es lo mismo que lo posible o lo probable. En otras palabras, lo más posible o probable no es siempre también lo mas previsible, ni vice versa. La que hace a una situación posible imprevisible no es su ser meramente poco probable, sino su ser racionalmente (y, por lo tanto, también responsablemente) insignificante, y ésta es una noción irreduciblemente normativa, no sólo descriptiva.

Desafortunadamente, la palabra “previsible”, en español, es ambigua.¹⁵ La noción relevante para entender lo que yo llamo situaciones **previsibles** no es la noción epistémica de ‘preveer’, sino la noción mas fuerte de ‘prevenir’. Prefiero el termino “previsible”, a “prevenible” porque este último se usa de manera común en el

¹⁵. Además, etimológicamente refiere al futuro, aunque nuestras metas pueden estar en el pasado o en el presente también. Agradezco a Catherine Elgin el haberme llamado la atención en este respecto.

sentido de “evitable”, pero lo que se busca evitar no es la situación previsible misma, sino las consecuencias negativas que podrían resultar de no estar preparado para ella. En otras palabras, uso “previsible” para referirme, no a lo que se puede preveer, sino lo que se debe prevenir. Prevenir no es sólo darse cuenta de que algo es posible – que es como comúnmente entendí la noción de ‘preveer’, sino responder de manera responsable ante este conocimiento, es decir, **prepararse** para que se dicha posibilidad.¹⁶ Así pues, una situación es mas previsible que otra, en el sentido relevante para entender la normatividad, no si es mas subjetivamente probable o si se da en un mundo posible mas cercano en algún sentido puramente descriptivo, sino si sería mas irresponsable no estar prevenido. Para ser mas precisos, una situación *A* es mas previsible que otra situación *B*, para un fin *F*, si una persona que persigue *F* y no previene la posibilidad de *A* (es decir, no hace lo suficiente para estar adecuadamente preparado para *A*) es más irresponsable que una persona que persigue *F* y no previene la posibilidad de *B*. Esto se deba a que la noción misma de “suerte” no es meramente descriptiva, sino también normativa. Tener suerte es haber obtenido algo que no se ha ganado. Esperamos que sea raro que quién esté bien preparado falle, es decir, esperamos que la suerte no sea la regla, sino la excepción. Pero lo que hace a los casos de suerte excepcionales no es su ser poco probables. En sentido estricto, lo que distingue a quién está preparado (en el sentido que aquí me interesa) de quien no lo está no es la pura **cantidad** de situaciones para las que está preparado, sino el **tipo** de situaciones para las que está preparado. Prepararse para un posible huracán

¹⁶. Colin Allen (2010) hace una distinción similar entre “predecir” el futuro y “anticiparse” a él; según Allen, predecir es saber cómo serán las cosas en el futuro, mientras que anticiparse es prepararse de manera adecuada a lo que podría pasar. Mi noción de *previsión* corresponde a su noción de *anticipación*.

Además, analogar ambas nociones es importante para reconocer que el fenómeno que me interesa en este libro no es un fenómeno exclusivamente humano, sino que está presente en muchos otros organismos, a saber, en cualquier animal que necesite controlar su comportamiento (que es algo aún más débil que tomar una decisión) en situaciones de incertidumbre. Carlos Montemayor (2014) y John Grecco (2010), entre otros, también han propuesto debilitar nuestra noción de agencia para extender estas explicaciones a animales no humanos. Aunque, en mi caso, mi interés es extenderlas también sistemas en los que la noción de agencia sería a lo más metafórica.

cuando se vive en zona de huracanes es racional y responsable en un sentido que no lo es prepararse para una invasión extraterrestre. Por eso, a la persona que se preparo para una invasión extraterrestre que no sucedió, pero no para un huracán previsible no le decimos que tuvo mala suerte, sino que fue negligente; mientras que a la persona que se preparo para un huracán previsible que nunca llegó, pero no para una invasión extraterrestre imprevisible no le juzgamos de negligente, sino a lo más, de tener mala suerte.

Desde esta perspectiva, decimos que estamos preparados (en el sentido relativo) para una situación si es posible garantizar que, si se da dicha situación, obtendré el resultado (si así lo deseamos), pero esto no significa que si se da una situación para la que no estamos preparados, es posible garantizar que **no** se va a obtener el resultado, sino que **no** es posible garantizar que **sí** se va a dar, es decir, se puede obtener o no, pero ya no depende de nosotros, sino de la suerte. En otras palabras, una vez que se da una situación para la que no estamos preparados, está fuera de nuestro control qué vaya a suceder – el mundo es comúnmente indiferente a nuestros deseos: se puede dar el resultado que deseamos o puede que no, dependiendo de nuestra suerte, buena o mala.

Comúnmente invocamos el estar preparados para algo (y nociones similares) cuando queré explicar cosas, especialmente cuando queré explicar el éxito o fracaso de una empresa. Decimos cosas como "La charla salió muy bien porque la preparé con cuidado", "El Canelo perdió la pelea porque aún no estaba listo para pelear con un campeón", etc. Esa es la noción que quiero recuperar aquí y, en este respecto, es importante notar que la relación entre el estar preparado para algo y el éxito o fracaso que ello explica no es una relación meramente causal. Efectivamente hay algún tipo de relación causal entre el estado resultante de mi preparación para X y X. Por ejemplo, para prepararme para mi próximo viaje a Ciudad Obregón, tendría que asegurarme de que mis maletas estén hechas con las cosas que necesitaré, que he hecho los arreglos necesarios para llegar a tiempo, comprando un boleto de avión, estando en el aeropuerto a tiempo, etc. Si efectivamente logro llegar a tiempo a mis compromisos en Ciudad Obregón, debe haber una relación causal entre algunos de los estados resultantes de mi preparación como el que tenga mi billete de avión a la hora de presentarme al aeropuerto y el estado de yo estar en Ciudad Obregón algunas horas después, pero aquí parece que lo causalmente eficaz no es mi preparación misma, sino algunos de los hechos físicos subyacentes. Además, probablemente habrá muchos hechos involucrados en mi preparación para X (en el sentido absoluto) que podrían ser ineficientes para causar

X, pero esenciales para estar listo para X (en el sentido absoluto). Por ejemplo, como veré mas adelante, podría sobrevivir a un huracán porque me preparé de manera adecuada sin que todo lo que hice (y necesitaba hacer) para estar listo para el huracán sea causalmente responsable de mi supervivencia real.

Por otro lado, la relación entre el estar preparado (en el sentido absoluto, pero no en el sentido relativo) y el éxito resultante tampoco puede ser una relación de fundamentación metafísica, si uno piensa – como es costumbre, pero creo yo que de manera equivocada (Barceló 2015) – que esta relación implica necesidad metafísica ya que, como he insistido a lo largo del curso, el vínculo entre la preparación (absoluta) y el éxito resultante es falible. Además, no parece que la preparación sea (ni siquiera parcialmente) constitutiva del éxito. Los fundamentos metafísicos de que mi equipo haya perdido su juego son todos los hechos que sucedieron durante el juego (en particular, que el balón cruzó nuestra línea de gol más veces de las que cruzó la de nuestro oponente), no lo que sucedió antes del partido. Por lo tanto, a menos que adopté una concepción falible de la fundamentación metafísica, como la que defiendo en (Barceló 2015) – parece haber un tipo distinto de vínculo que no es ni causal, ni metafísico (y presumiblemente tampoco conceptual ni lógico) entre el estar preparado (en el sentido absoluto) para algo y que hayamos logrado ese algo.

Dado que en la mayoría de los casos cuando uno está preparado (en el sentido absoluto) para algo, no se está completamente preparado (en el sentido relativo) para cualquier situación eventual que se pueda dar, es compatible con el estar preparado que no se logre el resultado esperado. Estos son los casos de mala suerte. Cuando creo, afirmo o, en general realizó una acción dirigida a un objetivo o que puede ser evaluada respecto a su resultado, puedo actuar preparado no preparado, pero dicha preparación rara vez va a garantizar de manera total el éxito de la empresa. De tal manera que, aunque esté preparado, mi empresa puede terminar en fracaso.

Por otro lado, y por supuesto, así como no es suficiente estar preparado para obtener el resultado deseado, tampoco es necesario. Si uno tiene buena suerte, puede obtener éxito sin haberse preparado, como el proverbial burro que tocó la flauta. Pero aún más, y esta es la moraleja principal del famoso argumento de Gettier, es posible estar preparado (en el sentido absoluto), que se de una situación para la cual uno no está preparado (en el sentido relativo) y sin embargo, se dé el resultado deseado. Es decir, así como se puede tener mala suerte, se puede tener buena suerte, aun cuando se esté preparado. Como había mencionado antes, ambos

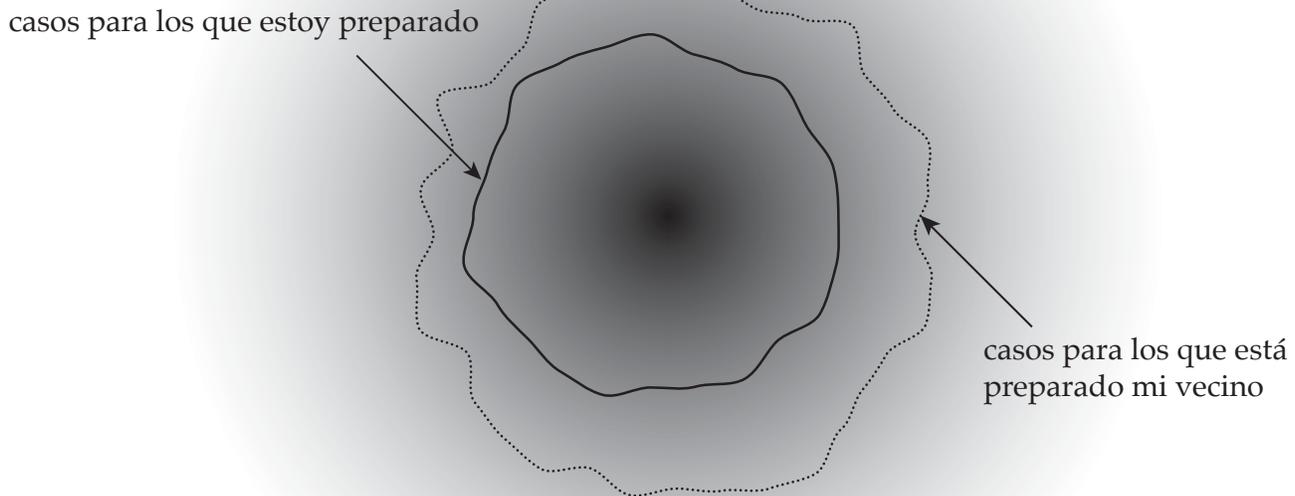
casos de suerte, buena y mala, tienen en común que resulten de que se dé una situación para la que uno no está preparado.

Pongamos un ejemplo para ilustrar esto: Supongamos que vivo en una zona de tormentas severas, donde es un riesgo presente la posibilidad de inundaciones. Si soy una persona responsable, contactaré a las autoridades adecuadas para enterarme si mi vivienda se encuentra en una zona de derrumbes o inundación repentina, además de informarme sobre los planes y señales de emergencia, incluyendo información sobre rutas de escape y localización de albergues. También realizaré regularmente simulacros de evacuación y tendré siempre a la mano provisiones de emergencia como agua, alimentos no-perecederos, equipo de primeros auxilios, etc. Cada uno de las cosas que haré para prepararme me ayudará para lidiar con algunas de las posibilidades que se pueden dar en caso de inundación. Por si se producen fisuras en las tuberías de gas, por ejemplo, habré cerrado las llaves maestras de abastecimiento de gas tan pronto como sepa del riesgo inminente de inundación, por ejemplo. Es posible que la tormenta no produzca fisuras en las tuberías de gas, pero es responsable estar preparado para dicha posibilidad y es por ello que cierro la llave maestra.

Entre mas me preparo para una posible inundación, mayores serán mis posibilidades de sobrevivir con el menor daño a mi propiedad y a la gente bajo mi responsabilidad. Y así como yo me preparo para la posible inundación, mis vecinos también se prepararán si son responsables. Es más, es posible que mi vecino se prepare **más** que yo. Si yo guardo comida para mantener a mi familia por cuatro o cinco días, es posible que él guarde comida para mantener a la suya durante dos semanas; si yo reviso la caducidad de mi extintor de incendios cada seis meses, él puede hacerlo cada tres, etc. Aun así, eso no hace que él sí esté preparado y yo no. Hay un umbral de 'preparacionez' tal que si uno lo pasa, no importa cuanto más se prepare, es correcto decir que uno ya está preparado.

Es posible usar el siguiente diagrama para explicar mejor esta idea.¹⁷ Imagínense este plano como el espacio de posibilidades de tal manera que cada punto de él representa una situación futura relevante ordenada

¹⁷. Tomo el diagrama de Miguel Ángel Fernández, aunque creo que no lo ha usado en su obra escrita, sino sólo en sus presentaciones orales.



casos para los que estoy preparado

casos para los que está
preparado mi vecino

de tal manera que entre más cercana al centro, más previsible es.¹⁸ Uno podría garantizar por completo el éxito sólo si estuviera preparado para cualquier situación futura posible. Sin embargo, el esfuerzo necesario para ello sería excesivo, así que es racional estar preparado sólo para las situaciones futuras más previsibles, es decir, para los puntos más cercanos al centro. Por supuesto, dónde dibujar la línea entre las posibilidades para las que vale la

¹⁸. Para aquellos lectores con inclinaciones más técnicas y formales, es importante notar que, dado que el fenómeno que me interesa modelar es hiperintensional (como ya mostró Yablo en 2013), estas posibilidades no pueden modelarse o identificarse con mundos posibles, sino que será necesario apelar a mundos parciales, situaciones o teorías sobre el mundo (Barwise 1983). Dado que mi texto no está dirigido a un público tan interesado en dichos detalles (importantes, sin duda), he preferido usar el término “posibilidades”, con todas sus ambigüedades.

pena prepararse y las que no es una cuestión imprecisa. Aun así, queda claro que no es necesario prepararse para las posibilidades más remotas para estar adecuadamente preparado. Hay un umbral vago, pero existe, entre las posibilidades que hay que tomar en cuenta y aquellas que no.

Supongamos entonces que yo he cruzado ese umbral, y aunque no estoy tan preparado como mi vecino, estoy adecuadamente preparado para una posible inundación. Esto significa que aunque mi riesgo es mayor que el de mi vecino, ambos he sido responsables en tanto ambos riesgos son razonables. En el diagrama, la línea oscura representa las situaciones futuras posibles para las que estoy preparado y la línea punteada las situaciones para las que mi vecino está preparado. Aunque la región cubierta por la de mi vecino es más grande que la mía, ambas cubren las posibilidades más previsibles. Pero esto significa también que habrá situaciones en las cuales, pese a estar ambos preparados en el sentido absoluto, yo no esté preparado y mi vecino sí. Esto puede significar que él y su familia sobrevivan el desastre sin mayor dificultad y mi familia sufra o ni siquiera sobreviva. El desastre puede ser tal que, por ejemplo, esté aislados no por tres o cuatro días, sino por dos semanas o más. En dicho caso, mis provisiones no serán suficientes, aunque las de mi vecino sí. Si eso sucede, uno estaría tentado a pensar que en realidad **no estaba** preparado, pero lo mejor sería decir que tuve muy mala suerte.

Cambié un poco el ejemplo e imaginé que la inundación es de tal magnitud que estamos aislados por más tiempo del que estábamos preparados, pero mi vecino decide compartir sus provisiones conmigo y mi familia y así nos ayuda a sobrevivir hasta que somos rescatados. En ese caso debé decir que estábamos preparados para la inundación y que la sobrevivimos, pero es claro que no es posible decir que la sobrevivimos **porque** estábamos preparados. En otras palabras, nos encontramos en un caso de Gettier. Estábamos preparados para las situaciones más normales que esperar, pero no para caso extraordinarios como el que de hecho se dio, y es por eso que si sobrevivimos o no lo hicé, no será porque estábamos preparados o porque no lo estábamos, sino por suerte: buena suerte si sobrevivimos, mala suerte si no lo hicé.

Trasladé lo que acabamos de decir sobre inundaciones al caso epistémico. Creer, había dicho ya, involucra siempre un riesgo: el riesgo de que lo que creamos sea falso, y lo epistémicamente responsable es hacer algo para evitar que esto suceda. Si me baso en evidencias y buenas razones, podré sentirme satisfecho en haber sido responsable con mis creencias aunque no pueda garantizar al cien por ciento que ellas sean verdaderas (no

puedo, por ejemplo, asegurarme que no soy un cerebro en una cubeta o que no estoy en alguna otra situación escéptica). Al creer, me comprometo a estar preparado para que el mundo sea como creo que es. Si el mundo es como es racional esperar que sea, entonces mis creencias serán verdaderas,¹⁹ pero no es posible garantizarlo.

Tener justificación para una creencia particular, digamos la creencia de que p , es estar preparado para que el mundo sea de ciertas maneras (las maneras en las que sería racional esperar que el mundo sea y por lo menos algunas maneras excepcionales), maneras en las que p es verdadero. Pero no es necesario estar preparado para cualquier manera que sea el mundo.²⁰ Ni el que tiene una creencia justificada ni el que la tiene sin justificación alguna están preparados para cualquier manera que sea el mundo, de tal manera que si el mundo es de alguna de estas maneras para las que no estamos preparados, no podré garantizar que nuestra creencia será verdadera. Es posible que sí lo sea, pero también es posible que no. De cualquier manera, no dependerá de nosotros, sino de la suerte. Si tenemos buena suerte, nuestra creencia será verdadera; si tenemos mala suerte, será falsa. La justificación no nos vacuna contra la suerte; a lo más, sólo disminuye la probabilidad de que dependamos de ella. Esto es lo que significa decir que la justificación reduce el riesgo de fallo. Tanto la buena como la mala suerte tienen en común que resulten de que se dé una situación para la que uno no está preparado.

Lo interesante es que esto se puede extender no sólo a otros casos donde se persigue un objetivo, sino a cualquier estado humano o acción que pueda evaluarse positiva o negativamente. Si creé, por ejemplo, que es bueno estar delgado, por ejemplo, es posible distinguir entre casos en los que se es afortunado ser delgado de aquellos en los que es un logro; si creé que la sopa que estamos comiendo es deliciosa, es posible distinguir entre casos en los que por suerte nos tocó una sopa deliciosa de casos en los que es el producto de la capacidad del cocinero de preparar una sopa deliciosa. Igualmente, es posible distinguir casos en los que “hasta el mejor

¹⁹. Esta relación entre la situación para la que estoy preparado y la verdad de lo que digo no es una inferencia (es decir, no infiero que el mundo es como creo que es de que el mundo es como es racional esperar que sea, sino que simplemente creo que el mundo es como creo que es y presupongo que el mundo es como es racional esperar que sea, pero esta presuposición no es un estado o acto mental separado) aunque bien puede haber una implicación lógica o metafísica aquí.

²⁰. Es mas, ni siquiera es necesario estar preparado para todas las maneras en las que el mundo podría ser y donde p es verdadero.

cazador se le escapa la liebre” de casos en los que la liebre se escapa precisamente porque somos malos cazadores. Y para cualquiera de estos valores es posible generar fácilmente casos de Gettier. Es posible imaginar casos en los que el buen cazador caza la liebre, pero no porque es buen cazador sino porque tuvo suerte. O casos en los que hicimos todo lo que sería razonable esperar de una persona responsable para ser delgados y aun así nuestra delgadez sea más el resultado de suerte que de nuestro esfuerzo por mantenernos delgados. Todos estos son casos Gettier.

| CREENCIA | VERDADERA | JUSTIFICADA | CONOCIMIENTO |
|-----------------------------------|--------------------------------|---|--------------------|
| ACCIÓN DIRIGIDA HACIA UN OBJETIVO | ALCANZAR EL OBJETIVO | Haber hecho lo suficiente para lograr el objetivo | LOGRAR EL OBJETIVO |
| EMPRESA | ÉXITO | PREPARACIÓN | LOGRO |
| ACCIÓN MORALMENTE EVALUABLE | BUENAS CONSECUENCIAS (NO DAÑO) | CUIDADO (NO NEGLIGENCIA) | |
| Aserción | verdad | justificación | conocimiento |

Aunque todos estos son espacios en los cuales es posible aplicar el esquema que aquí he dibujado, a los filósofos le interesan sólo ciertos tipos de acciones y estados y sólo ciertos tipos de normas y valores. Por ejemplo, a la epistemología suele interesarle la creencia y tomar a la verdad como criterio fundamental para evaluarla. A la filosofía del lenguaje, en contraste, suelen interesarle los actos de aserción, aunque también suelen tomar a la verdad como criterio fundamental para evaluar dichos actos. Cuando afirmo que el mundo es de cierta manera, me comprometo a que, a menos que suceda algo poco previsible, entonces será como afirmo que es. En ambos ámbitos este modelo se aplica de manera directa. Sin embargo, también es posible aplicarla a otros campos del lenguaje de interés filosófico. Por ejemplo, como ya mencionaba en la sección anterior, también es posible entender el uso que se da de los predicados o conceptos bajo este esquema. Idealmente, poseer un concepto debería de implicar ser capaz de determinar para cualquier objeto si cae bajo su extensión o no. Sin embargo, es racional usar un concepto aun cuando no seamos capaces de determinar su extensión para cualquier objeto

posible, siempre y cuando esté preparados para hacerlo en los casos más previsibles. Es por ello que nuestros conceptos son vagos. Así pues, es posible concebir la posesión de un concepto como el estar preparado para usarlo y a la vaguedad como una manifestación previsible de la falibilidad de nuestras capacidades. En el próximo capítulo, desarrollaré algunas de estas aplicaciones esperando que mi propuesta adquiera mayor definición a través de este tipo de ejemplos.

IV. Aplicaciones y Ejemplos

A. Casos Gettier en Matemáticas

Una vez que he extraído las moralejas de la falibilidad de nuestros medios y del argumento de Gettier, la pregunta relevante para la filosofía de la matemática es si es posible explotar estos nuevos espacios teóricos para renovar su agenda epistemológica. ¿Es posible entender mejor el conocimiento matemático si empezamos por plantearnos que el quehacer matemático no ha dado los resultados que ha dado por accidente no suerte? Si es así ¿cuales son las habilidades o capacidades relevantes? ¿Probar teoremas? ¿Sistematizar intuiciones? Dadas las obvias limitaciones de espacio y tiempo, en este texto me enfocaré sólo a la primera de estas putativas habilidades: el probar teoremas. Además, el abordar la cuestión de si el probar teoremas es un método infalible de producción de conocimiento es muy relevante para la cuestión de si efectivamente todos nuestros mecanismos y métodos de producción de conocimiento son falibles, como sostiene el falibilista, o existe por lo menos alguno que sea infalible. A fin de cuentas, casi siempre que se aborda este problema suele salir a colación el caso de las pruebas matemáticas formales como un posible contraejemplo al falibilismo.

Para comenzar, vale la pena empezar preguntándose si la noción de “prueba” relevante para la epistemología de las matemáticas es una noción trascendente o instrumental, es decir, si la regla de que debé creer sólo en proposiciones matemáticas probadas es una regla trascendente que sólo puede servirnos para definir una de nuestras metas epistémicas, o si es una regla instrumental que puede servirnos para guiar nuestro actuar como matemáticos. La cuestión es fundamental y no es sorprendente que haya sido una de las preocupaciones centrales de filósofos y matemáticos durante el período conocido como “crisis de los fundamentos” a finales del siglo XIX y principios del siglo XX. Esta preocupación es central al proyecto epistemológico de, por ejemplo, David Hilbert. Su programa tenía por objetivo central asegurarse de que la noción de “prueba” operativa en matemáticas fuera, al mismo tiempo, infalible y que puede caracterizarse en términos no-trascendentes, es decir, que el método de pruebas requiriera para su implementación sólo de capacidades cognitivas básicas que estuvieran a disposición de cualquier matemático. Por eso Hilbert restringió

su noción de ‘prueba’ sólo a aquellas que pertenecieran a sistemas formales cuyas reglas de deducción y axiomas pudieran definirse de manera recursiva con recursos finitos (a decir verdad, la noción misma de recursividad tenía como objetivo capturar de manera rigurosa nuestra noción de calculable por cualquier persona). Desafortunadamente, gracias al trabajo de Gödel y otros, hoy sabemos que es imposible producir una noción de prueba que satisfaga los tres requisitos que buscaba Hilbert, es decir, que fuera infalible, instrumental y nos permitiera conocer todas las proposiciones matemáticas verdaderas. Esto significa que la noción de prueba operativa en matemáticas es una noción trascendente o falible, es decir, que cuando en matemáticas se prueba algo, de hecho, no se trabaja con una noción de prueba formal en el sentido de Hilbert.

Hay, sin embargo, otras razones para pensar que la noción formal de ‘prueba’ es trascendente. Por principio de cuentas, hablamos de “probar un teorema”, no cuando se hace el esfuerzo, sino cuando se alcanza obtener la prueba. Si el resultado no es una prueba, no decimos que hé probado un teorema, sino sólo que lo hé intentado. En otras palabras, para que podamos llamar a una acción “probar un teorema” es necesario que su resultado sea una prueba. Cabe por lo tanto preguntar si es suficiente que el resultado de una actividad sea una prueba para hablar de probar o es necesario algo más, algo así como el empleo de una habilidad. En otras palabras ¿es posible probar algo por accidente (en el sentido relevante a este ensayo)? Supongamos que un pésimo estudiante, que no ha estudiado en clase ni ha desarrollado la habilidad de probar teoremas en, digamos, geometría diferencial, desesperado decide improvisar en su examen una secuencia de fórmulas al azar las cuales, afortunadamente para él, terminan formando una prueba perfecta del teorema de Stokes. ¿Diríamos que el estudiante sabe que la integral de una función f en el intervalo $[a, b]$ puede ser calculada por medio de una antiderivada F de f ? Seguramente no. Parece que algo le falta. Lo mismo diríamos de alguien quién, sin el menor conocimiento de geometría diferencial, se propusiera aprenderse de memoria la secuencia de fórmulas que forman una prueba de dicho teorema. Es posible decir que este último personaje es capaz de producir a voluntad pruebas del teorema de Stokes. Sin embargo, me parece que en este caso tampoco podríamos decir que sabe que dicho teorema es verdadero.

¿Qué consecuencias epistemológicas podríamos sacar de estos ejemplos? Por un lado, parece ser que para probar un teorema no es suficiente producir una prueba. ¿Significa esto que el probar es un *logro* en el sentido arriba caracterizado? La respuesta no es muy obvia. En el primer ejemplo, podríamos decir que precisamente la razón por la cual lo que hizo el estudiante no es probar el teorema de Stokes es por que su prueba no fue el resultado de ninguna habilidad. No tenemos la menor seguridad de que, en otras circunstancias similares, el estudiante pueda volver a ser capaz de producir dicha prueba. Sin embargo, en el segundo caso sí tenemos a alguien ejerciendo una habilidad para producir la prueba. Nuestro segundo personaje sí puede volver a probar el teorema a voluntad (en las circunstancias adecuadas, obviamente). El resultado no es accidental, y por lo tanto, debería poder ser considerado un logro, no una mera obtención de un resultado positivo.

Tal parece que enfrentamos un horrible dilema: o bien tenemos que rechazar la hipótesis de que para probar un teorema basta lograr probar un teorema de manera no accidental, o rechazamos el presupuesto de que lo que es capaz de producir el segundo de nuestros personajes es un prueba propiamente dicha. Aunque sea una teoría del error, este segundo cuerno del dilema no es tan ad-hoc como podría parecer. Después de todo, no es raro escuchar a filósofos decir que las secuencias de fórmulas o en general el tipo de cosas que se pueden escribir en exámenes y aprender de memoria sin haber entendido **no** son pruebas propiamente dichas. Para fundamentar esta concepción de la prueba matemática, revisé un poco más cierta concepción tradicional de la prueba como justificación epistémica. Tradicionalmente, la justificación que otorgan las pruebas a nuestras creencias en la verdad de los teoremas probados se considera una justificación indirecta o inferencial. Comúnmente se piensa que el problema epistemológico fundamental para las matemáticas es el de la justificación de nuestras creencias en los axiomas de nuestras teorías, ya que se piensa que nuestras creencias en el resto de los teoremas están justificadas por las pruebas. Bajo esta concepción, las pruebas son un tipo de inferencia, es decir, procesos psicológicos. Por eso, no es posible decir que el estudiante de nuestro ejemplo ni quién aprendió de memoria las fórmulas efectivamente probaron un teorema, ya que la inferencia no sucedió en la mente de ninguno de ellos.

Ahora bien, si las pruebas son procesos inferenciales, entonces regresamos a nuestra pregunta anterior ¿es suficiente haber obtenido el resultado buscado (la inferencia válida) para decir que se probó el teorema? Es decir ¿es posible probar un teorema por accidente? ¿Es posible encontrar escenarios parecidos a los anteriores, pero no sobre la prueba escrita, sino sobre la prueba psicológica? Supongamos ahora un segundo estudiante, uno que sí sabe un poco de geometría diferencial, pero que como ser falible que es, aún se le van algunas inferencias, es más, a veces se le van algunas inferencias lógicas básicas. Por ejemplo, a veces se confunde acerca de cómo funciona la implicación material y a veces cree que es válido inferir la negación del consecuente de la negación del antecedente. Además, también llega a confundir cuál es el antecedente y cuál el consecuente. Supongamos ahora que a la hora que se propone probar el teorema de Stokes realiza todos los pasos a la perfección, excepto en un paso en el cual tiene ya probados un condicional y la negación de su consecuente. Por error, identifica mal el antecedente y el consecuente del condicional, por lo que cree haber probado el condicional y la negación de su antecedente. En ese momento también se confunde sobre cómo aplicar el *Modus Tollens* y cree que éste valida inferir la negación del consecuente a partir de la negación del antecedente. El resultado, por supuesto, es una inferencia válida, un *Modus Ponens*. Si no vuelve a cometer un error, el estudiante habrá realizado puras inferencias válidas y así inferido válidamente el teorema de las premisas relevantes. ¿Diríamos por lo tanto que el estudiante probó el teorema? Admito que mis intuiciones a este respecto no son tan firmes y estables como las de los ejemplos anteriores, pero el epistemólogo tradicional tendrá que aceptar que sí, que el estudiante ha probado el teorema. Según la epistemología tradicional, no importa si la prueba es resultado del azar o del esfuerzo, efectuarla es suficiente para tener conocimiento. En consecuencia, el estudiante de nuestro ejemplo efectuó una inferencia válida, y aunque creía que su inferencia era válida, no lo **sabía**. Carece de un conocimiento de segundo orden sobre la fuente de su conocimiento. Cree correctamente que su inferencia fue válida, pero no lo sabe. No sabe que ha probado lo que cree correctamente haber probado.

Hay por lo menos otros problemas sustanciales para quienes quieran defender este tipo de epistemología. Por un lado, tenemos el problema de que si una prueba es una inferencia (o una serie de inferencias), hay pruebas tan largas que la inferencia sería humanamente imposible, y parece tener poco sentido hablar de una inferencia que no se puede realizar y que, sin embargo, tenga algún valor epistémico.

Una manera de tratar de rescatar la intuición de que el estudiante de nuestro ejemplo no sabe el teorema que ha probado, aunque tenga una prueba correcta de él, y que pruebas como la del mapa de cuatro colores tienen valor epistémico, es pedir de un sujeto, para saber la verdad de un teorema matemático, no sólo que posea o haya realizado una prueba, sino que también **sepa** que dicha prueba sea correcta. Sin embargo, esto debilitaría la motivación central detrás de esta opción, la cual era establecer una analogía fuerte entre la epistemología de las matemáticas y el resto de la epistemología. Por el contrario, lo que se establecería sería una separación infundada entre el papel epistémico de la prueba matemática y la inferencia en el resto del conocimiento, donde es ampliamente aceptado que -- excepto en casos extraordinarios -- la inferencia transmite conocimiento.

B. Casos Gettier Morales

Ya he argumentado que, dadas las fuertes similitudes entre los ámbitos de lo moral y lo epistémico (Barceló 2012), debí esperar encontrar casos de Gettier en el caso de juicios morales, es decir, casos en los que se actúe con malas intenciones, y en general, el agente haga todo lo que esté en su disposición para producir un mal,²¹ el mal que se buscaba se de hecho como efecto de los esfuerzos y actos del agente y, sin embargo, haya algún sentido o perspectiva desde la cual el acto no es moralmente reprobable. Esta tercera perspectiva, he argumentado, es la perspectiva del vínculo adecuado entre las intenciones (y los preparativos, esfuerzos, etc. del agente) y los resultados de la acción. Así como un caso Gettier en epistemología es uno donde uno tiene una creencia verdadera justificada que no es conocimiento porque la justificación y la verdad no están adecuadamente vinculadas, un caso Gettier moral sería uno donde se tienen malas intenciones, un mal resultado, y sin embargo, no estaríamos dispuestos a calificar la acción de mala en un tercer sentido (aún no explorado adecuadamente en la ética) porque el resultado no está adecuadamente ligado a la intención. En otras palabras,

²¹. Por economía, de ahora en adelante, cuando hable de intenciones me referiré no sólo a las intenciones, sino también a los esfuerzos, habilidades, capacidades, etc. del agente relevantes para determinar que el agente hizo todo lo que estaba de su parte para lograr sus nefastos fines.

si mi hipótesis central de este curso es correcta, es decir, si todos los ámbitos normativos tienen la misma estructura, es de esperar encontrar casos de Gettier en todos los ámbitos normativos, incluido el moral.

Recientemente, Carolina Sartorio (2015) ha publicado un interesante artículo sobre lo que ella llama una "nueva forma" de suerte moral. Su caso proviene de la literatura sobre el papel de la causalidad en el derecho, y ha llegado a ser conocido como el acertijo del "viajero en el desierto". La versión de Christopher Taylor y Daniel Dennett es la siguiente:

“Todos en el puesto de avanzada de la Legión Extranjera Francesa odian a Fred y lo quieren muerto. La noche anterior a que Fred realice una caminata por el desierto, Tom envenena el agua de su cantimplora. Luego, Dick, sin saber de la intervención de Tom, vierte el agua (envenenada) y la reemplaza con arena. Por último, Harry llega y perfora la cantimplora, para que el "agua" se escape lentamente. Más tarde, Fred se despierta y emprende su viaje, provisto de su cantimplora. Demasiado tarde se da cuenta de que su cantimplora está casi vacía, pero además, lo que queda es arena, no agua, ni siquiera agua envenenada. Fred muere de sed.” (Christopher Taylor and Daniel Dennett 2000. Mi traducción)²²

Este se trata de un caso de Gettier moral. Lo que hizo Tom es malo desde la perspectiva de sus intenciones y medios (el análogo moral de la justificación epistémica), es decir, Tom ha hecho lo suficiente de lo que le correspondía como para estar justificado en esperar que Fred muriera de sed como resultado de sus acciones. Así como la justificación es falible, esta manera de actuar mal no garantiza plenamente que resulte el mal esperado. Además, Fred murió de sed, lo que significa que se alcanzó el objetivo de Tom al envenenar el agua de Fred. Sin embargo, no diríamos que Tom logró matar a Fred. La muerte de Fred no fue un logro de Tom (y mutatis

²². Everybody in the French Foreign Legion outpost hates Fred and wants him dead. During the night before Fred's trek across the desert, Tom poisons the water in his canteen. Then, Dick, not knowing of Tom's intervention, pours out the (poisoned) water and replaces it with sand. Finally, Harry comes along and pokes holes in the canteen, so that the "water" will slowly run out. Later, Fred awakens and sets out on his trek, provisioned with his canteen. Too late he finds his canteen is nearly empty, but besides, what remains is sand, not water, not even poisoned water. Fred dies of thirst.

mutandi para Harry o Dick) porque no hubo un vínculo apropiado entre las acciones de Tom y el resultado deseado, es decir, porque Fred no murió de sed debido a que Tom envenenó a su agua, sino por suerte.

Carolina Sartorio tiene razón al señalar que este es un caso de suerte moral en el sentido de Nagel, pero substancialmente diferente de los que suelen presentarse en la literatura (donde una acción es mal intencionada pero tiene buenas consecuencias o viceversa), porque aquí la acción es mal intencionada, produce un mal, pero no es mala en el sentido de que el mal que se produce no es un logro basado en las malas intenciones del agente. Por lo tanto, es posible mantener la hipótesis central de este curso de que todo campo normativo involucra (por lo menos) estas tres perspectivas de evaluación distintas que corresponden, en el caso epistémico, a la justificación, la verdad y el conocimiento. En el caso moral, si tengo razón en mi interpretación del acertijo del viajero por el desierto, la acción de Tom es justificadamente malvada (actuó mal), tiene un mal resultado pero no hay un vínculo adecuado entre la justificación maligna y el mal resultante, por lo que no estamos hablando de un acto malo pleno, pues a fin de cuentas Tom no logró matar a Fred.

G. Hugo Padilla sobre Gettier

Hugo Padilla (1971) ha sugerido que parte de la razón por la cual durante tanto tiempo se nos escaparon casos como los de Gettier es lingüística. En particular se debe a que, en muchos contextos normales, cuando alguien afirma una frase de la forma “ P y Q ” (para dos enunciados P y Q), los escuchas entienden que la persona ha tratado de comunicar algo más que la mera conjunción de P y Q , ya sea que Q sucedió después de P o que fue un efecto o consecuencia de P . Es por ello, piensa Padilla, que muchos filósofos les cuesta trabajo separar las hipótesis de que el conocimiento es creencia verdadera y justificada de la hipótesis de que el conocimiento es creencia que es verdadera *porque* está justificada, y así darse cuenta de que los ejemplos con los que Gettier ilustra su argumento son contraejemplos de la primera hipótesis pero no de la segunda.

El punto, sin embargo, es más general; nos muestra la importancia de ser muy cuidadosos al presentar casos tipo de Gettier, tanto dentro como fuera del campo epistemológico. Consideré un ejemplo como el de los posibles ataques en Berlín. Si presentamos los hechos del caso así:

1. Hubo un intento de ataque con camiones.
2. Estábamos preparados para un ataque con camiones.
3. El intento de ataque falló.

Sería decir algo verdadero, pero engañoso pues sugiere fuertemente que el fallo se debió a los esfuerzos de las fuerzas anti-terroristas de Berlín, en vez de la incompetencia de los atacantes, lo cual es falso. Pero esto no se debe a ninguna cuestión profunda sobre la naturaleza de la normatividad o en particular sobre la falibilidad de las medidas adoptadas en Berlín, sino mas bien a imprecisiones en nuestro lenguaje. Siguiendo la sugerencia de Padilla, si mi propuesta es correcta, aquí hay una especie de falacia de equivocación, la cual se hace aun mas manifiesta cuando se hace explícito donde sería el ataque, por ejemplo, así:

- 1'. Hubo un intento de ataque con camiones (en la zona despoblada X de Berlín).
- 2'. Estábamos preparados para un ataque con camiones (en la zona poblada Y distinta de X).
- 3'. El intento de ataque falló.

Así presentado el ejemplo, desaparece la sugerencia de que el intento falló gracias a que estábamos preparados y nos hace preguntarnos ¿porqué falló el intento de ataque en X? En otras palabras, mientras que la descripción en (1), (2) y (3) parece completa, la de (1'), (2') y (3') es claramente incompleta.

Lo mismo se aplica a los otros ejemplos que he ya visto:

1. Hubo una inundación.
2. Estábamos preparados para una inundación.
3. Sobrevivimos.

Presentar así los hechos es decir cosas verdaderas pero engañosas pues se sugiere que la razón por la cual sobrevivimos es porque estábamos preparados. Sin embargo, una descripción mas adecuada – no porque la primera no fuera verdadera, sino porque escondía elementos relevantes para darse cuenta de lo que realmente sucedió sería:

1'. Hubo una inundación (de diez días).

2'. Estábamos preparados para una inundación (de cuatro o cinco días).

3'. Sobrevivimos.

Presentar así los hechos nos hace preguntarnos inmediatamente ¿cómo sobrevivimos si no estábamos preparados para una inundación de diez días? En otras palabras, presentar así las cosas muestra que lo que falta es precisamente una conexión entre lo que de hecho sucedió y *para qué* estábamos preparados y que esta conexión se puede perder cuando la posibilidad que se actualiza en los enunciados de tipo (1) – y, por lo tanto, en la que se funden los hechos descritos en los enunciados de tipo (3) – no pertenecen a las posibilidades de las que dependen los hechos descritos en los enunciados de tipo (2).

Si mi análisis es correcto, lo mismo debe suceder en los ejemplos que usa Gettier. Tomé el primer contraejemplo. En él, Smith y Jones han solicitado el mismo empleo. El gerente de la empresa le ha asegurado a Smith que Jones se quedará con el empleo. Luego, Smith sabe que el número de monedas que Jones tiene en el bolsillo son diez, porque las ha contado (de manera competente). Finalmente, sin que Smith mismo lo sepa, él también tiene diez monedas en el bolsillo y será quien se quedará con el empleo. Una vez mas, presentar este caso de la siguiente manera es tramposo:

1. El hombre que se quedará con el empleo tiene diez monedas en su bolsillo.
2. Smith cree justificadamente que el hombre que se quedará con el empleo tiene diez monedas en su bolsillo.
3. La creencia de Smith de que el hombre que se quedará con el empleo tiene diez monedas en su bolsillo es verdadera.

Es engañoso, no porque sea falso – tanto (1) como (2) y (3) son verdaderas en la situación descrita por Gettier – sino porque sugiere fuertemente que el que Smith haya logrado tener una creencia verdadera es consecuencia de su justificación (o, si se quiere, de aquello en lo que descansa su justificación). Una vez mas, sostiene Padilla, tenemos una equivocación que se manifiesta con claridad una vez que describimos el caso con mayor detalle, así:

1. El hombre que se quedará con el empleo (Smith) tiene diez monedas en su bolsillo.

2. Smith cree justificadamente que el hombre que se quedará con el empleo (Jones) tiene diez monedas en su bolsillo.
3. La creencia de Smith de que el hombre que se quedará con el empleo tiene diez monedas en su bolsillo es verdadera.

Esto deja claro que la situación relevante en (1) – el que Smith se quedará con el empleo y tiene diez monedas en la bolsa – no pertenece al conjunto de situaciones previstas por Smith en su justificación (2) – el contar las monedas en la bolsa de Jones, por ejemplo, no le da ninguna información sobre las monedas en su propia bolsa (la justificación de Smith es insensible a los hechos que hacen a su creencia verdadera). Esto fortalece la tesis que he venido defendiendo de que todos estos casos son estructuralmente idénticos.

D. El Lenguaje como Tecnología

Antes de terminar de estudiar la falibilidad, quisiera presentar un ejemplo de importancia filosófica en el que, como espero mostrar a lo largo del curso, es útil hablar de falibilidad y no suele hacerse: las palabras. No seré el primero en sostener que el lenguaje es una tecnología y que las palabras son sus herramientas, pero me parece que las consecuencias filosóficas de esta afirmación aún no han sido exploradas del todo. Para los objetivos de este curso, es útil pensar así a las palabras para poder entender los límites de la semántica tradicional y cómo adoptar un anti-infalibilismo nos permitiría superarlos.



(C) 2014 RSTEVENS :: DIESELSWEETIES.COM

Existe una manera tradicional, y de sentido común, de pensar el lenguaje, entre otras cosas, como un medio de representación, es decir, como una manera de representar cómo es el mundo. Los predicados, por ejemplo, nos sirven para describir maneras en que las cosas pueden ser o no ser. Hay cosas que son verdes y otras que no lo son y la función del predicado “verde” es captar esta diferencia. El significado de “verde”, pues, es la propiedad que tienen las cosas verdes y que las distingue de las que no lo son. En consecuencia, aprender el significado de “verde”, es adquirir la capacidad de usarla para marcar dicha diferencia.

En oposición a esta manera de concebir el lenguaje, filósofos del lenguaje ordinario como Wittgenstein, Strawson, Brandom o Ezcurdia nos invitan a pensar al lenguaje como una tecnología y a las palabras como sus herramientas. Como tales, son instrumentos que usamos para diferentes fines, entre ellos, el de representar cómo es el mundo. Pero rechazan que esto se logre meramente a través de una correspondencia directa entre palabra y realidad de tal manera que, por ejemplo, a cada predicado le corresponda una propiedad y a cada nombre un objeto. En su lugar, proponen pensar al lenguaje en su uso. Esto significa, entre otras cosas, que cuando decimos que usamos al lenguaje para comunicar y representar al mundo lo que queremos decir es que comunicar y representar al mundo son cosas que los humanos hacemos con la ayuda del lenguaje, es decir, que las palabras son las herramientas que usamos para realizar esta tarea y como tales hay que pensarlas. En este sentido, en vez de decir que los nombres refieren a objetos, lo más correcto sería decir que somos nosotros los que nos referimos a objetos usando sus nombres. De la misma manera, el significado de un predicado como “verde”, por ejemplo, no es la presunta propiedad que tienen las cosas verdes y que las distingue de las que no lo son, sino el uso que le damos a dicha palabra para distinguir ciertas cosas de otras. Como Charles Travis (1985) de manera célebre ya señaló, probablemente no haya ninguna propiedad que tengan en común una fruta cuando es verde, un semáforo cuando está en verde ni el planeta tierra que haga adecuado que a todas ellas les apliqué el adjetivo “verde”.

Esto no significa, por supuesto, que las palabras no tengan significado y que podamos usarlas de cualquier manera. Si alguien me pregunta cómo me pareció la última película que vi, no puedo usar el adjetivo “verde” para comunicar que me pareció muy bien lograda y fiel al libro en el que está basada. Para explicar porque es posible usar las palabras para lo que de hecho las usamos es esencial apelar a sus significados convencionales –el giro pragmático en filosofía del lenguaje no propone lanzar todos los diccionarios a la hoguera – pero este significado convencional no puede por sí mismo fijar de manera completa, para cada posible circunstancia de uso, a qué objetos nos es posible referir y a cuales no. A esto es a lo que me referiré cuando diga que las palabras como herramientas no son instrumentos infalibles de referencia.

Ahora bien, para entender cabalmente qué significa sostener que las palabras son herramientas, uno debe tener claro qué son y cómo funcionan las herramientas. Por ejemplo, si decimos que conocer el significado de una palabra es saber usarla como herramienta, esto significa que para poder desarrollar una teoría del significado de las palabras necesitamos saber qué implica saber usar una herramienta en general. Tomé un ejemplo sencillo de herramienta: un martillo. Supongamos que alguien quiere enseñarle a otra persona a usar un martillo. Probablemente lo que haga sea détrarle como se agarra, como se toma el clavo, y como se levanta para tomar fuerza y ejercerla sobre la cabeza del clavo para que este se inserte en la superficie sobre la que queré clavarlo, etc. Sin embargo, imaginé que el aprendiz le pregunte a su maestro: “Muy bien, veo lo que hiciste ahí, pero no me has dicho exactamente cuanta fuerza debo aplicar ni cuanto exactamente debo levantarlo, o siquiera me has dicho de qué tamaño deben ser los clavos.” Es posible que uno trate de resolver la preocupación del aprendiz explicando como es necesaria la suficiente fuerza para vencer la resistencia de la superficie a clavar o como diferentes tamaños y tipos de claves se usan para diferentes materiales y con diferentes funciones. Pero, ¿qué sucedería si el aprendiz insistiera en que fuéramos mas precisos? Uno podría decir, bueno, si insertas el clavo de manera perpendicular al grano de la madera, debes meterlo por lo menos 12 veces el valor de su diámetro, Por ejemplo, si usas un clavo de 4mm de diámetro, debes introducirlo en la madera por lo

menos 4.8 centímetros. Si no penetra lo suficiente, la pieza se puede zafar, mientras que si lo introduces demasiado, la madera se puede partir.

Hasta aquí todo suena muy sensato. Sin embargo, es posible imaginar un extraño aprendiz que quisiera insistir en preguntar exactamente cuanto debe entrar del clavo en la madera. “Me hablas de ‘demasiado’ y ‘lo suficiente’, pero ¿exactamente cuanto es eso? – puede insistir nuestro extraño aprendiz – ¿debo introducirlo exactamente 12 veces su diámetro?” En este caso es claro que uno nunca va a poder dar una respuesta absolutamente exacta a su aprendiz y va a llegar un momento en que deba reconocer que estas reglas son aproximadas y que o bien no hay tal cosa como la distancia exacta que el clavo deba introducirse a la manera o que si la hay uno no la conoce. Sería aun mas extraño, pero concebible que al escuchar esto el aprendiz concluya entonces que, en realidad, el maestro no sabe como usar el martillo, en tanto no sabe cuanto – es decir, exactamente cuanto – debe introducirse el clavo en la madera. Aun peor sería si el aprendiz concluyera que el problema es con el martillo (y el clavo) y sostuviera que es inservible en tanto no se conoce su uso y que es necesario desarrollar una nueva herramienta, una especie de martillo perfecto, para el cual sí haya parámetros exactos y conocidos de cuanta fuerza debe ejercerse, de qué tamaño debe ser el clavo, cuanto debe penetrar cada material, etc.

Pero por supuesto que no hay ningún problema con el martillo ni con nuestro conocimiento de su uso. Sabemos usar el martillo aun cuando este conocimiento no pueda expresarse sino con reglas aproximadas. Cuando aprendí a usar el martillo, como casi cualquier herramienta, no lo hice a través de reglas exactas, sino a través de una combinación de detración de uso, adquisición de reglas aproximadas, ensayo y error.

Ahora bien, para los filósofos del lenguaje ordinario, la comunicación lingüística es como la carpintería y las palabras son como los clavos y el martillo, nuestras herramientas. Para saber usarlas no necesitamos reglas precisas. Por ejemplo, no necesitamos saber exactamente a qué temperatura debe estar un objeto para saber si se le aplica la palabra “caliente” o no. No necesitamos saber exactamente qué tan lejos debe estar un objeto para poder decirle “ese” en vez de tener que decirle “este”, para decir que está “allá” en vez de “aquí”. Y si hubiera un objeto en esa area gris para la cual no tenemos claro si

es apropiado decirle “aquí” o “allá”, basta con buscar otra manera de identificarlo, ya que para eso tenemos otras palabras y otros mecanismos. El que haya una zona gris entre el “aquí” y el “allá” no denota ningún problema con estos términos, ningún defecto que los haga inservibles, aunque sí denota cierta limitación a su uso. Denota que no es posible determinar, en cualquier situación, para cualquier objeto espacial, si es adecuado describir su ubicación usando “allá” o no (y lo mismo sucede con “aquí”), es decir, denota que como herramienta no es infalible sino que tiene limitaciones. Pero conocer este tipo de limitaciones es parte del saber usar dichos términos, aun cuando este conocimiento sea él mismo vago – es decir, aunque tampoco podamos decir exactamente a qué distancia de nosotros se encuentra esa línea gris. Sería igual de necio que el aprendiz de carpintero de nuestra historia anterior el filósofo que insistiera en reglas exactas para el uso de nuestras palabras. Probablemente es una cuestión de control la que lleva a algunos filósofos del lenguaje a tomar una actitud frente al lenguaje similar a la que toma nuestro aprendiz de carpintero frente al clavo y al martillo.

Pero lo importante para nuestros objetivos de pensar a las palabras como herramientas es que nos permite entender mejor su aspecto normativo. Cuando decimos que el significado de una palabra es su uso, obviamente no queremos decir que es posible usar cualquier palabra cómo queramos y que cómo sea que la usé, la habré usado bien, en tanto ése es su significado. Como cualquier otra herramienta, las palabras pueden usarse bien o mal, y por lo general, no basta la intención de ningún usuario individual para hacerla hacer algo para o que no tiene la capacidad. Así como mi propia voluntad no es suficiente para convertir esta extensión eléctrica en un sacapuntas, así tampoco puedo decidir un día usar la palabra “máscara” para referirme al aroma de la tierra recién mojada tras un chubasco. En tanto el uso de las palabras es comunitario, como bien insistió tanto Wittgenstein en su obra tardía, se necesitan por lo menos dos para usar una palabra. Pero es sólo en esa situación que se pone a prueba su efectividad y eficiencia como herramienta comunicativa. En tanto herramientas de coordinación humana, las palabras tienen significados públicos y externos, es decir, al igual que cualquier otra herramienta, las palabras obtienen su significado en respuesta a presiones externas, materiales y sociales en un proceso complejo, dinámico y fragmentado. Y aunque los significados son externos en este sentido, también son

internalizados por sus usuarios en tanto les sirven para alcanzar sus fines de comunicación y coordinación.

En el modelo tradicional, saber qué significa una palabra significa saber en qué situación es correcto usarla, de tal manera que si, por ejemplo, llamo “verde” a algo que no es verde, este error puede deberse a una de dos cosas: a que pienso equivocadamente (o quiero hacer creer a los demás que pienso) que es verde o a que me equivoco sobre lo que significa “verde”. En otras palabras, si yo sé que algo no es verde y aun así le llamo “verde” esto se debe a que no sé bien qué significa la palabra verde. En el modelo de palabras como herramientas, en contraste, saber qué significa una palabra significa saber en qué situación es correcto usarla, de tal manera que si uso la palabras “verde” por ejemplo, para tratar de describir un objeto, algo habra salido mal si no logro que mi escucha identifique la propiedad a la que me refiero; pero esto no significa necesariamente que el error haya sido resultado de mi falta de pericia en el uso del lenguaje, es decir, no necesariamente debe evidenciar que no conozco el significado de verde o que no soy capaz de detectar correctamente el color del objeto del que hablo. Tal vez puede ser que tuve mala suerte y cosas que razonables asumí sobre el contexto no eran tal y como las pensaba. En otras palabras, surge una tercera fuente de posible error que no proviene, ni del significado, ni del referente. Proviene de la suerte.

El significado de una palabra garantiza la referencia exitosa sólo en condiciones normales, es decir racionalmente previsibles. El que estamos en una situación racionalmente previsible de comunicación es lo que comúnmente llamamos, en filosofía del lenguaje y lingüística, una presuposición. En ausencia de la presuposición, la referencia escapa de nuestras manos y se convierte en un asunto de suerte. Esto significa que es posible que haya errores de comunicación aun en el intercambio comunicativo de dos personas bien informadas y hábiles en su uso del lenguaje.

V. Deontologismo y Consecuencialismo

“Las personas normales —personas que aparte de ésto no tienen ningún signo de desorden o falta de control de los principios morales humanos fundamentales— hacen cosas malas *todo el tiempo*. Uno podría escribir millones de páginas detallando todas las veces en que los actos malvados fueron perpetrados, instigados o no resistidos por personas que, en cualquier otro aspecto, eran perfectamente normales. Es seguro decir, a cierta aproximación, que todos nosotros —realmente quiero decir esto; realmente quiero decir que usted y su familia y todos los que usted ama— podría, en un contexto histórico diferente, haber tenido esclavos o haber contribuido al Holocausto o por lo menos podría haber decidido que no valía la pena oponerse a estos grotescos crímenes. Porque esa es la condición humana: no es posible acceder de manera fácil a una perspectiva amplia de la moralidad y la empatía. Hací lo que la gente que nos rodea está haciendo, lo que nuestra cultura está haciendo.”

Singal (2017)²³

A. El Valor de Lograr Algo

Habíamos señalado ya que reconocer la falibilidad de nuestros logros nos obliga a distinguir entre lograr algo y obtenerlo por mera suerte. Esto nos ha llevado a dos importantes preguntas filosóficamente sustanciales: la cuestión *fundacional* de porqué valoramos más un logro (debido al esfuerzo propio o al ejercicio de alguna capacidad) que el obtener lo que deseamos de manera accidental o por buena suerte, mientras que castigamos

²³. “Normal people — people who otherwise have no signs of derangement or a lack of a grip on basic human moral principles — do evil stuff *all the time*. One could write millions of pages detailing all the times when evil acts were perpetrated, abetted, or not resisted by people who were, in every other respect, perfectly normal. It’s safe to say, to a certain approximation, that all of us — I really mean this; I really mean you and your family and everyone you love — could, in a different historical context, have been a slaver or a Holocaust-perpetrator or at the very least decided it wasn’t worth the trouble to contest these grotesque crimes. Because that’s the human condition: We don’t have easy access to a zoomed-out view of morality and empathy. We do what the people around us are doing, what our culture is doing.” Mi traducción, itálicas en el original.

igual los casos en los que el mal es resultado de mala suerte que cuando no lo es (Martin & Cushman 2015), y la cuestión *descriptiva* de explicar en qué consiste exactamente el que lograr un objetivo se deba al ejercicio de alguna capacidad o sea el resultado directo de hacer un esfuerzo para alcanzarlo. El capítulo pasado estuvo dedicado a la pregunta fundacional, mientras que éste lo dedicaré a dar respuesta a la cuestión descriptiva, es decir, explicar exactamente qué tipo de relación debe haber entre el esfuerzo o ejercicio de una capacidad y la obtención del objetivo para decir que alcanzar dicho objetivo fue un logro y no un afortunado accidente. Nótese que para ello no basta que el resultado deseado sea causado por la posesión de la capacidad o la realización del esfuerzo. Regresé al ejemplo del que trabaja para comprarse un coche. Supongamos que para no gastarse el dinero que iba guardando para comprar su coche, nuestro personaje abre una cuenta bancaria de esas que automáticamente te registra a una lotería. Si se gana el coche en dicha lotería, habrá una relación causal entre su esfuerzo por ahorrar para comprarse el coche y la obtención de dicho coche. Sin embargo, es claro que su obtención del coche no se debió de manera directa a su esfuerzo o al ejercicio de alguna capacidad propia. En otras palabras, no fue un logro personal, sino producto de la suerte. Adopté, como es costumbre en la literatura, la convención de llamar “consecuencia” a la relación que hay entre un esfuerzo virtuoso por lograr algo y el logro que se obtiene de él (Berker 2013). El problema fundacional, entonces, es preguntarse porque valoramos más lograr algo como consecuencia de nuestra esfuerzo que por suerte, y el problema descriptivo es explicar precisamente en qué consiste esta relación de consecuencia.



Respecto a la pregunta fundacional, en un extremo tenemos lo que llamaré el “**consecuencialismo radical**” – que no es sino lo que Hilary Kornblith llama “consecuencialismo burdo” y Bortolotti (2016) llama “universalismo” – según el cual es un error (muy generalizado) el pensar que el hacer el esfuerzo para lograr algo o tener la capacidad de lograrlo tengan algún valor, cuando lo único que tiene realmente valor es aquello que se logra (Harman 1999, Gardner 2007, Littlejohn *en prensa*). Otras palabras, el esfuerzo que no termina en éxito no

vale nada. Por lo tanto, para este tipo de consecuencialismo, no debé valorar diferente un logro de un accidente afortunado.

Mas generalizado es un tipo mas moderado de consecuencialismo, según el cual sólo las metas tienen un valor fundamental, y los logros y medios confiables sólo tienen valor de manera derivada en tanto nos acercan a esas metas valiosas. Este tipo de consecuencialismo también sostiene que es un error muy generalizado el pensar que el hacer el esfuerzo para lograr algo o tener la capacidad de lograrlo tengan algún valor independiente del valor de aquello que logran; sin embargo, no piensan que no tenga ningún valor. Aunque si valoramos la posesión o el ejercicio de una capacidad no es sino porque valoramos obtener aquello que es posible alcanzar con el ejercicio de dicha capacidad. En otras palabras, y tal y como lo caracteriza Berker (2013):

Todas las teorías consecuencialistas [moderadas] sostienen que (i) hay ciertos estados de cosas que son inherentemente buenos, y otros estados de cosas que son inherentemente malos, y (ii) todas las demás nociones normativas están determinadas por lo bien que conducen hacia ellos de manera directa o indirecta, en virtud de estar adecuadamente conectados con alguna otra entidad que conduzca directamente hacia ellos. Si aplicamos esta idea general al ámbito ético y tomamos al bienestar individual como el valor meta (sin tomar en cuenta cómo se distribuye ese bienestar), lo que resulta es el utilitarismo. Si aplicamos esta idea general al reino epistémico y tomamos a la creencia verdadera como el valor meta, lo que resulta es el confiabilismo. Lo que importa epistémicamente, para los confiabilistas, son las consecuencias (en términos de alcanzar creencias verdaderas y evitar creencias falsas) de formar creencias de una manera determinada. (Berker

En el otro extremo se encuentran filósofos de inclinación Kantiana (pienso en Christine Korsgaard 2000, por ejemplo, pero también Bortolotti 2016, entre otros) para quienes la diferencia fundamental entre el logro y el accidente afortunado es que sólo el primero puede adjudicarse normativamente al sujeto. En otras palabras, para este tipo de *deontologistas* – por llamarlos de alguna manera –, hay una diferencia sustancial entre que algo sea malo/bueno y que alguien haya actuado mal/bien, es decir, que algo sea correcto/incorrecto. Si alcanzamos nuestras metas esto es algo bueno, pero no habré actuado bien (es decir, de manera correcta) a menos que dicha meta haya sido lograda en base a nuestros esfuerzos y cuidados. Por eso no elogiamos a quienes ganan la lotería (aunque, como veré a detalle mas adelante, sí asignamos responsabilidad e imputamos castigos a quienes desafortunadamente actúan o causan males sin que estos sean intencionales o el resultado de negligencia por su parte). Para entender mejor este tipo de deontologismo, vale la pena distinguir dos sentidos en los que evaluamos positivamente algo. Por ejemplo, cuando queré ir de picnic, es posible decir que es bueno que el clima esté agradable, pero al hacerlo no estamos haciendo un juicio moral genuino, ya que no estamos evaluando ningún acto ni ningún agente. No hay nada eminentemente humano en la bondad del clima, aun cuando su bondad dependa de nuestras necesidades y preferencias humanas y que también sea muy humano nuestro aprecio. Sin embargo, cuando decimos que es bueno que alguien haya traído, repelente de mosquitos al picnic, lo que

²⁴. “All consequentialist theories hold that (i) there are certain states of affairs that are inherently good, and other states of affairs that are inherently bad, and (ii) all other normative notions under the theory’s purview are determined by how well the objects of assessment conduce toward or promote those states of affairs (either by directly conducing toward them, or by indirectly conducing toward them, in virtue of being suitably connected to some other entity that directly conduces toward them). If we apply this general idea to the ethical realm and take the value to-be-promoted to be individual well-being (with no regard to how that well-being is distributed), what results is utilitarianism. If we apply this general idea to the epistemic realm and take the value to-be-promoted to be true belief and the avoidance of false belief, what results is reliabilism. What matters epistemically, for reliabilists, are the consequences (in terms of attaining true beliefs and avoiding false beliefs) of forming beliefs in a certain way.”

estamos evaluando si es un acto genuino ya que estamos adjudicando la bondad de un estado conveniente – el que conté con repelente de mosquitos en el picnic – a una persona – a quién lo trajo. Para que este trasladar la bondad de la consecuencia al agente sea correcto, sostienen los deontologistas, es necesario que la consecuencia sea un logro del agente, es decir, es necesario que sea resultado de sus esfuerzos, capacidades, etc. y no de la suerte. Al trasladar esta bondad al agente, se vuelven apropiadas nuevas respuestas que antes no estaban disponibles. Por ejemplo, es posible felicitar o reprender al agente por sus la bondad o maldad de sus actos (y las consecuencias de sus actos), y si nosotros somos el agente, es posible sentirnos orgullosos o avergonzados de ellos. Por eso no significa lo mismo decir que el clima es bueno para un picnic (es decir, es propicio para un picnic) a decir que el clima es el correcto para un picnic (es decir, es correcto irse de picnic en un día así); lo que estamos evaluando en cada caso es algo muy distinto, y sólo en el segundo caso lo que estamos evaluando es un acto.²⁵

Finalmente, hay una posición aun mas **radical** en oposición al consecuencialista en la cual los valores con los cuales evaluamos a los medios son completamente independientes de los valores con los que evaluamos los fines, de tal manera que un mecanismo, regla o método puede ser valioso aun si **no** nos acerca a nuestros fines, es decir, si **no** reduce el riesgo de falla. Estoy pensando en filósofos como Hilary Kornblith, Catherine Elgin y Sally

²⁵. Dado que a mí me interesa extender este modelo normativo a casos donde lo que quere evaluar no son actos o estados de agentes, es necesario relativizar este principio también a entidades que no sean agentes. Así pues, aun cuando hablamos de una herramienta, por ejemplo, es posible distinguir entre las bondades que es posible apreciar de las consecuencias positivas de usar dicha herramienta, de las que podemos ascribir a la herramienta misma por estar basadas en su funcionamiento correcto. Sobre las similitudes y relaciones entre la normatividad apropiada para evaluar a agentes humanos y la normatividad relevante para evaluar herramientas, véase Franssen (2006). Como mencionaba anteriormente, también hay muchos intentos por extenderlo a organismos que tienen, a lo más, una agencia mínima (Montemayor 2014).

En otras palabras, aun cuando hay razones filosóficas para distinguir el tipo de normatividad que sólo es apropiado aplicar cuando hay agencia genuina y robusta (Fernández 2013), también hay ventajas filosóficas en concentrarnos en los aspectos mas abstractos que se dan independientemente de si estamos tratando con casos de agencia, agencia mínima o no.

Haslanger para quienes nuestras obligaciones y nuestro sentido de ‘cuidado’ se han construido a partir de contingencias históricas y/o evolutivas que no dependen todas de nuestros fines. En este sentido, por lo tanto, el mero llamarlos medios es ya confundente, pues denota un lugar secundario y derivado con respecto a los fines. Así por ejemplo, dado que nuestras capacidades cognitivas son el resultado de un proceso evolutivo que ha sido el resultado de muchas contingencias históricas que se han dado a lo largo de nuestra historia biológica, es completamente posible que ellas funcionen a la perfección sin llevar de manera confiable a la verdad. Extendiendo esta idea a otros ámbitos, es posible decir que los métodos que adoptamos para alcanzar nuestros fines son muchas veces el resultado de contingencias históricas, psicológicas, etc., muchas de las cuales no responden a qué tanto dichos métodos nos acercan o no a nuestros fines. En palabras de Cecelia Watson, “las tecnologías ... están profundamente insertas en los contextos en los que fueron desarrolladas y, por tanto, están sujetas a tendencias y gustos.” (Watson 2012: 651, n.7. Traducción mía)²⁶ Por ejemplo, es una regla común entre los bibliotecarios de Estados Unidos poner libros de calidad en la sección de libros para niños para formar lectores mas exigentes, aunque no hay muchos datos empíricos rigurosos que sustenten dicha regla (y hay razones para pensar que no sirve de nada para mejorar la calidad de los libros que los niños que asisten a la biblioteca de hecho leen, Edmonds 1987); esto significa que un bibliotecario responsable en Estados Unidos pondrá libros de calidad en los libreros de la sección para niños aunque es muy posible que esto no contribuya en nada a su fin, a saber, formar lectores mas exigentes. Esto se debe a contingencias históricas y sesgos cognitivos que han llevado a los bibliotecarios Estadounidenses a adoptar esta regla (y a convencerse de su valor) independientemente de su confiabilidad o contribución al fin al que se supone está dirigida. Sin embargo, eso no hace inadecuado, en ausencia de información extra (por ejemplo, de que de hecho dicha práctica no sirve para formar lectores más exigentes), el juzgar negativamente al bibliotecario que no siga dicha regla. Aunque se deba a contingencias históricas y sesgos cognitivos, esta regla sigue siendo parte de lo que es ser un bibliotecario responsable en el

²⁶. “[Technologies] are deeply imbedded in the contexts in which they were developed, and thus subject to trend and taste.” Traducción mía.

contexto actual de los Estados Unidos.²⁷

Así pues, a la pregunta de ¿porqué valoramos mas los logros que (el mero alcanzar) las metas?, consecuencialistas y deontologistas dan respuestas radicalmente distintas. Para el consecuencialista radical, como habíamos mencionado, no debé valorar diferente un logro de un accidente afortunado. Ahora bien, la razón por la cual *parece* que valoramos los medios confiables y los logros es porque en realidad nos dan evidencia, derrotable pero confiable, de que se darán los fines, los cuales son los que realmente tienen valor. En otras palabras, lo único que realmente nos interesa es alcanzar nuestros fines; sin embargo, la mayoría de las veces no sabemos si obtendremos o no lo que deseamos, a lo único que tenemos acceso es a nuestros medios. Por eso buscamos en ellos información, indicios sobre lo que nos interesa en realidad, es decir, si obtendré nuestros fines o no. En ausencia de conocimiento sobre los fines, apelamos a información sobre la confiabilidad de los medios para hacer nuestros juicios evaluativos tentativos. Sin embargo, esos juicios siguen siendo acerca de los fines, no acerca de la confiabilidad de los medios. La clave aquí es distinguir claramente entre que algo sea bueno o malo y que algo esté conectado evidencialmente con el que algo bueno o malo haya pasado o vaya a pasar. Por ejemplo, si no sabemos si alguien actuará bien o mal, información sobre sus intenciones nos puede servir para realizar un juicio tentativo sobre sus actos ya que de sus intenciones es posible inferir, derrotablemente, el tipo de actos que es posible esperar de él o ella. Pero al hacerlo, no estamos evaluando sus intenciones, sino que estamos usando sus intenciones como información evidencial para juzgar lo que realmente nos interesa: sus actos. De la misma manera, antes de saber si lo que alguien dice es verdad o falsedad, es posible servirnos de sus razones para adelantarnos en nuestro juicio sobre lo que dirá y anticipar la verdad de lo que dirá, pero otra vez no estaré valorando sus razones, sino usándolas como evidencia de lo que nos interesa, a saber, la verdad de lo que dice.

²⁷. Por supuesto, el consecuencialista ha tratado de dar cuenta de este tipo de fenómenos argumentando que lo que estos aparentes contra-ejemplos muestran es que nuestros actos y decisiones no persiguen *un* sólo fin y que en casos como éste tendríamos que tomar en cuenta el contexto social y los fines que de él emanan (Bar-Eli 2009). Así, por ejemplo, los bibliotecarios siguen esta regla no porque les ayude a alcanzar su *presunto* fin de formar lectores mas exigentes sino porque les ayuda a alcanzar fines sociales como el de mantener su estatus dentro del grupo. De esta manera, casos como este dejan de ser contra-ejemplos a la hipótesis consecuencialista.

En consecuencia, una vez que sabemos si lo que dice es verdadero o falso, sus razones se vuelven insustanciales. Igualmente, para el consecuencialista radical, para poner otro ejemplo, parece que valoramos el dinero por sí mismo, cuando en realidad lo que nos interesa es lo que es posible comprar con él y si parece que valoramos el dinero, en realidad, es porque al poseerlo anticipamos el valor de lo que conseguiré a través de él. En todos los casos, los medios no tienen valor alguno, pero sí pueden darnos información sobre si es posible esperar o no aquello que sí es de valor, es decir, las metas.

Para el consecuencialista moderado, en contraste, los medios sí pueden tener algún valor, pero éste no es completamente independiente del valor de los fines, sino que se deriva de él. Esto significa que los medios confiables son valiosos precisamente porque nos acercan a nuestros fines de manera confiable. Pero este valor, aunque derivado, es genuino, en el sentido de que un medio confiable es valioso aun en esos casos en los que, de hecho, no nos lleva a alcanzar nuestros fines. Así pues, al buen cazador que se le escapa la liebre le es posible reconocer el valor instrumental de su habilidad como cazador, aun en aquellos casos en que dicha habilidad haya sido infructuosa. En consecuencia, el alcanzar nuestras metas por mera suerte tiene la desventaja, sobre el lograr nuestras metas a partir de nuestros esfuerzos, cuidados, etc., de ser un éxito endeble, frágil, uno que no es posible esperar se repita; lo que explica su inferioridad frente al logro, el cual es robusto, repetible y confiable.

Finalmente, el deontologista propone una distinción fuerte en el tipo de valor que es posible adjudicar a las metas y el que le es posible adjudicar a los logros.²⁸ Las metas tienen un valor similar al que tienen otras cosas que nos pueden agradar o convenir, como la lluvia u otras cosas buenas de la vida que no siempre es posible adjudicar a algún agente; mientras que los logros tienen un valor eminentemente humano. Alcanzar una de nuestras metas, no a través de la manifestación de nuestras capacidades, habilidades o esfuerzos, sino por buena suerte es bueno en el mismo sentido en que es bueno el buen clima o el haber nacido en una familia de dinero y al igual que ellos no es posible adjudicarlo a nadie; mientras que lograr nuestras metas es bueno en otro sentido, uno que nos atañe directamente como seres humanos, uno que dice algo sobre nosotros, sobre nuestro carácter.

²⁸. En realidad, como he visto, no son dos tipos de valor distintos, sino tres: uno que corresponde a las metas, otro a los medios y uno tercero a los logros.

Esto explica porque preferimos los logros a (el mero alcanzar) las metas: porque los primeros son virtuosos por partida doble: por un lado, son cosas buenas, y por el otro, son actos buenos, es decir, es posible rastrear su bondad a los actos correctos del agente. En otras palabras, aparecen como valiosos dentro de dos regímenes distintos: el de las cosas y el de los actos. Sin embargo, de los dos tipos de valores, sólo los segundos – el de los actos – es genuinamente normativo. A la realidad no es posible normarla, aunque preferiríamos que fuera de cierta manera en vez de otra. Sólomente hay normas ahí donde hay acción. La realidad no obedece o desobedece normas, sólo es cómo es.

Para el deontologista, una vez que le otorgamos relevancia normativa a consecuencias imprevisibles, si lo que hace es correcto o incorrecto se vuelve algo inestable y, tal vez, hasta incognoscible. En ciencia ficción, un tropo muy común, conocido como “La Mariposa de la Fatalidad”, nos dice que si pudiéramos regresar en el tiempo, lo que sabemos ahora sobre las consecuencias de lo que de hecho hicimos no sería suficiente para garantizar mejores consecuencias. Tal vez el ejemplo reciente mas obvio de este tropo es la película de 2004, *The Butterfly Effect*, escrita y dirigida por Eric Bress y J. Mackye Gruber, y agradezco inmensamente a Angel Andrés Cantú Jiménez el señalar su relevancia para el debate entre deontologistas y consecuencialistas.

La mariposa de la fatalidad suele verse como una consecuencia pesimista del así-llamado “Efecto Mariposa”, según el cual, dado un marco de tiempo lo suficientemente amplio, para cualquier acto siempre existirán consecuencias que sean sustanciales e impredecibles. Esto significa que si pensamos que todas las consecuencias de nuestros actos son relevantes para su evaluación, entonces no estará en nuestras manos el actuar bien o mal, pues siempre será posible, para cualquiera de nuestros actos, el que tenga una consecuencia imprevisible suficientemente importante como para afectar su valor. En otras palabras, no importa qué tan cuidadosos y virtuosos seamos en nuestras intenciones y actos, éstos siempre podrán llegar a tener en algún momento una consecuencia lo suficientemente negativa como para pensar que hubiera sido mejor actuar de manera distinta. Es preferible, por lo tanto, concluyen los deontologistas, restringirnos a lo que está bajo el control y es previsible para el agente, a la hora de evaluar sus actos.

En resumen, el debate entre deontologistas y consecuencialistas corresponde a dos maneras diferentes de

concebir a los medios y su relación con los fines. Desde la perspectiva deontológica, una persona está justificada en tomar un riesgo si, ya sea de manera consciente o automática, toma las medidas necesarias para actuar con el debido cuidado. Desde la perspectiva consecuencialista, una persona está justificada en tomar un riesgo si, dadas las probabilidades de obtención de cada resultado, el cálculo de costos y beneficios favorecen la toma del riesgo. Aunque diferentes, suele considerarse que ambos criterios están íntimamente relacionados: que la razón por la cual es bueno tomar cuidado en lo que hacemos, es porque así reducimos objetivamente el riesgo de no lograr nuestros objetivos. Igualmente, se espera que una explicación consecuencialista de lo que hace tomar un riesgo racional o no capture nuestras intuiciones acerca de lo que es racional esperar de un agente cuidadoso. Además, una vez que introducimos la distinción tripartita entre alcanzar un fin, hacer lo correcto para conseguirlo (es decir, estar justificados en esperar el fin), y lograr un fin (es decir, alcanzarlo *porque* hicimos lo correcto para conseguirlo), tiene sentido preguntarse cual de estas tres nociones es fundamental. Tradicionalmente, los consecuencialistas piensan que alcanzar la meta es lo que único tiene valor en sentido fundamental, mientras que el logro y la justificación sólo lo tienen en sentido derivado, a lo más; mientras que los deontológicos piensan que el valor de la justificación no se deriva del valor de la meta, sino que es valioso en sí mismo.²⁹

B. Un ejemplo matemático

Consideraré ahora un caso presentado por Selim Berker en su artículo de (2013), y que pretende poner en problemas a los consecuencialistas, especialmente a aquellos epistemólogos que sostienen que para que una

²⁹. Por supuesto, cabe la posibilidad que no haya nada detrás del debate entre deontologistas y consecuencialistas mas que preferencias personales de los propios filósofos y filósofas que develan algo de su carácter, como ha sugerido Gilles Deleuze (2001). Según el filósofo francés, el consecuencialista muestra rasgos sádicos, mientras que el deontologista muestra rasgos masoquistas, en tanto el primero concibe a la normatividad como fundamentada en algo externo y el segundo la concibe como emanando de nuestras propias prácticas y situaciones sociales.

creencia sea verdadera es suficiente que, además de ser verdadera, sea generada por un proceso que tiene a generar mas creencias verdaderas que falsas:³⁰

Supongamos que lo siguiente es verdadero sobre mí: siempre que contemplo si un número natural dado es primo, formo la creencia de que no lo es. “¿Es 25 un número primo? No, no lo es.” “¿Es 604 un número primo? No, no lo es.” “¿Es el 7 un número primo? No, no lo es.” También estipulé que este es el único proceso cognitivo a través del cual me formo creencias sobre números primos (Soy un tipo ingenuo). Puesto que la proporción de números primos a número compuestos menores que n se aproxima a 0 cuando n se acerca al infinito, mi proceso de formación de creencias tiende a producir una razón de creencias verdaderas a falsas que se aproxima a 1. Por lo tanto, los confiabilistas ... se ven obligados a decir que, debido a que mi proceso de formación de creencias es casi perfectamente fiable, cualquier creencia formada sobre su base está justificada. ¡Pero eso es una locura! Cuando creo una creencia de que 7 no es primo, simplemente no es correcto decir que, aunque esa creencia es falsa, es epistemológicamente redimida por la verdad de las otras creencias que se formarían a través del proceso que la condujo. Así, su teoría produce el veredicto epistémico equivocado.³¹

Aún cuando el problema está formulado explícitamente contra un tipo particular de consecuencialista, no es difícil ver cómo se puede generalizar a otro tipo de consecuencialismos epistémicos, y dado que mi modelo pretende recuperar del consecuencialismo epistémico la idea de que es una virtud de nuestros mecanismos el acercarnos a nuestros fines – y que, por lo tanto, una virtud de nuestros mecanismos epistémicos es el acercarnos a

³⁰. Como he señalado con anterioridad, para los objetivos de este trabajo la distinción entre procesos y mecanismos no es importante, por eso no es importante tampoco la cuestión de si lo que debe ser confiable es el mecanismo o el proceso que genera la creencia.

³¹. “Suppose the following is true of me: whenever I contemplate whether a given natural number is prime, I form a belief that it is not. “Is 25 prime? No, it is not.” “Is 604 prime? No, it is not.” “Is 7 prime? No, it is not.” Let us also stipulate that this is the only cognitive process by which I form beliefs about the primeness of natural numbers. (I’m a simpleminded kind of guy.) Since the ratio of prime to composite numbers less than n approaches 0 as n approaches infinity, my belief-forming process tends to yield a ratio of true to false beliefs that approaches 1. Therefore process reliabilists are forced to say that, because my belief-forming process is almost perfectly reliable, any belief formed on its basis is justified. But that’s crazy! When I form a belief that 7 is not prime, it is simply not correct to say that, although that belief is false, it is epistemically redeemed by the truth of the other beliefs which would be formed via the process that led to it. Thus even process reliabilists countenance cross-propositional trade-offs, and where they do, it seems their theory yields the wrong epistemic verdict.” Mi traducción.

nuestros fines epistémicos, a saber, la generación y mantenimiento de creencias verdaderas –, vale la pena explorar cómo trataría un problema como este.

En este respecto, un punto central a destacar del ejemplo de Berker es que es uno de esos confusos experimentos mentales en los que se nos pide suponer que una persona – en este caso, una persona concreta, Berker mismo – es completamente distinto, en aspectos epistémicamente fundamentales, a las personas reales (y en particular, sustancialmente distinto del verdadero Berker de carne y hueso) para luego pedirnos nuestras intuiciones sobre el estatus epistémico de unas creencias que las personas comúnmente no tienen, producidas por un proceso cognitivo que ninguna persona posee. El caso es tan abigarradamente contrafáctico, que es natural cuestionar su relevancia para entender nuestras prácticas de evaluación epistémica reales. Sin embargo, me parece que el ejemplo de Barker **sí** es muy ilustrativo de ciertos aspectos importantes de nuestras prácticas evaluativas, pero no exactamente por las razones que Berker señala. El ejemplo de Barker, me parece, nos puede ayudar a arrojar luz sobre, por ejemplo, porqué nuestras prácticas epistémicas sobre juicios matemáticos – y juicios científicos en general – son tan distintas de aquellas que versan sobre juicios cotidianos; y también sobre la diferencia entre evaluar actos y estados epistémicos humanos y otros estados y procesos representacionales análogos que, por lo tanto, también podrían caer dentro del dominio de aplicación de nuestros valores epistémicos.

Imaginé el mismo caso aplicado, no a un proceso inferencial humano, sino a uno artificial:

Supongamos que Roberta desarrolla un nuevo algoritmo para realizar una tarea dada; este nuevo algoritmo es más rápido que los algoritmos anteriores gracias a su nuevo sistema de procesamiento de datos, el cual asume que los datos a procesar, al ser codificados numéricamente, nunca corresponden a un número primo. En otras palabras, el algoritmo recibe un paquete de datos codificado en forma de un número, y aun cuando no sabe nada sobre qué número es, asume de entrada que no es primo, es decir, lo registra y procesa como si no fuera primo. En otras palabras, para el algoritmo, el número no es primo. Si el dato de entrada es 2500098405061834, para el algoritmo, no es primo. Si el dato de entrada es 1124098488569631, para el algoritmo, no es primo. Si el dato de entrada es 0000000000000007, para el algoritmo, tampoco es primo. También estipulé que el algoritmo no cuenta con ningún otro mecanismo o subrutina para verificar si un número es primo o no (es un algoritmo simple). Puesto que la proporción de números primos a número compuestos menores que n se aproxima a 0 cuando n se acerca al infinito, el algoritmo

tiende a producir una razón de resultados correctos a incorrectos que se aproxima a 1. ¿Cómo debé evaluar dicho algoritmo?

Aún cuando el nuevo ejemplo es análogo al de Berker en los aspectos relevantes, nuestras intuiciones sobre las virtudes del algoritmo difieren sustancialmente de las que teníamos en el caso anterior. Me parece que en este caso es muy claro que nuestro juicio dependerá, entre otras cosas, de cual es el riesgo real que se corre porque el algoritmo no funcione, ¿estamos hablando de un algoritmo que, de fallar, podría costarle la vida a alguien, por ejemplo, un algoritmo clave dentro de un aparato médico? ¿o estamos hablando de una simple subrutina que a lo más podría ‘crashear’ una aplicación de entretenimiento de nuestro teléfono móvil? (Mackenzie 2001) También dependerá de qué tanto se usará el algoritmo, ¿será un algoritmo que correrá continuamente de tal manera que, es posible esperar casi con certeza que en algún momento fallará? ¿o es un algoritmo de emergencia que sólo se usará de vez en cuando? y, finalmente, dependerá también en lo valioso que sea el tiempo que se ahorrará con el uso del nuevo algoritmo. Si bien es cierto que, por lo general, más rápido suele ser mejor; también es cierto que en muchos casos la velocidad no es importante o hasta puede ser contraproducente. Si el algoritmo no recibe información con suficiente velocidad o si el resto del sistema no es capaz de procesar resultados con la velocidad con la que el mecanismo los produce, por ejemplo, serían dos circunstancias en las que un aumento de velocidad no representa una ventaja sustancial.

Esto significa que, una vez que pasamos de hablar de creencias y procesos cognitivos humanos y tratamos de aplicar el mismo esquema a un mecanismo artificial, las intuiciones consecuencialistas regresan con fuerza. ¿Muestra esto que, por lo tanto, sí hay una diferencia importante entre la manera que evaluamos el funcionamiento de un artefacto que el de un humano, de tal manera que el consecuencialismo puede ser adecuado en el caso de las máquinas, peor no en el de las personas? Creo que no. Más que una diferencia importante entre mecanismos artificiales y humanos, el ejemplo de Berker sirve para reforzar esta semejanza en la que he insistido a lo largo del curso: que nuestras prácticas normativas son sensibles a la disponibilidad de recursos. Y aquí es donde el hecho de que el ejemplo proviene de las matemáticas es importante.

A la hora de evaluar un experimento mental es esencial preguntarnos si la descripción del caso es lo suficientemente completa respecto a los rasgos que son relevantes según las diferentes hipótesis en disputa

(Dowell, manuscrito). En tanto mi hipótesis sostiene que nuestras prácticas de evaluación – y, por lo tanto, nuestros reportes de conocimiento – dependen sustancialmente de los recursos disponibles al agente en el contexto (entre otros factores contextuales), ésta sostiene que cualquier descripción de una situación que no incluya esta información estará incompleta y que el que la escuche tratará de llenar esta información con los valores por defecto en el contexto. En el caso del experimento mental de Berker, por ejemplo, no sabemos, entre otras cosas, por qué está Berker pensando en números primos, qué tan urgente o importante es tener información correcta sobre esto, etc. factores que, sostengo, juegan un papel importante en nuestras evaluaciones de conocimiento. Esto hace que los que interpretamos el texto de Berker necesitamos completar la descripción con los valores estándar que se asumen por defecto en nuestro contexto. En este caso, tratamos de identificar la manera *normal* por la cual alguien necesitaría decidir sobre un número natural si debe creer que es primo o que no es primo. En otras palabras, de inmediato nos mové a un contexto matemático.

La ciencia en general, y las matemáticas en particular, son prácticas epistémicas bastante *sui-generis*. A diferencia de un gran número de nuestras preocupaciones cotidianas, y con contadas excepciones, las preguntas que se plantean en la ciencia carecen de la menor urgencia. Excepto por ciertos casos de investigación médica, climática, etc. no hay ninguna prisa real en encontrar respuesta a la mayoría de las preguntas fundamentales de la ciencia. Sí, por supuesto, como cualquier quehacer humano, el quehacer científico se realiza en un contexto de fechas límites, proyectos con plazos determinados, evaluaciones de resultados a plazos regulares, etc. Aún cuando hay importantes factores prácticos que determinan la aceptación de, por ejemplo, ciertos axiomas y conjeturas matemáticas que son valiosos para poder obtener otros resultados importantes, por lo general, no es el caso que necesité resolver un problema matemático *ya* y debamos adecuar nuestros estándares de aceptaciones a nuestras limitaciones de tiempo. En otras palabras, cuando se trata de proposiciones matemáticas, el tiempo es un factor desdeñable. Al eliminar restricciones de tiempo, muchos otros recursos materiales dejan también de ser significativos a la hora de determinar estándares epistémicos. Aun si le dedicamos pocos recursos matemáticos a, digamos, la investigación geométrica; los recursos acumulados durante más de veinte siglos de investigación geométrica resultan enormes. Esto significa que las matemáticas, y probablemente esto sea algo que se pueda

generalizar al resto de la ciencia, son empresas epistémicas de recursos superlativos. Esto tiene profundas consecuencias epistémicas.

En este curso he sostenido, al igual que muchos otros filósofos, que nuestros estándares epistémicos (y no epistémicos también) son sensible a la disponibilidad de recursos. En particular, cuantas (y cuales) posibilidades debemos tomar en cuenta para determinar si una creencia está justificada o no dependerá de con cuales recursos conté para realizar nuestra pesquisa. Por lo general, si tenemos mas recursos, es nuestra responsabilidad considerar mas posibilidades; si tenemos pocos, al contrario, es justificable considerar unas cuantas menos. Esto significa que aquellas investigaciones a las que dediqué mas recursos, tendrán estándares epistémicos mas altos. De ahí que las matemáticas y otras ciencias tengan estándares epistémicos tan altos.

Por eso es que, cuando nos preguntamos sobre las circunstancias en las cuales alguien está siquiera justificado en creer una proposición matemática como que 7 es un número primo, automáticamente pensamos en estándares muy altos, casi de certeza, estándares que nuestras prácticas actuales de prueba y aprendizaje sí sobrepasan. No en balde, en nuestro contexto socio-histórico, la manera canónica de obtener conocimiento matemático es a través de las práctica sociales de aprender y probar resultados matemáticos. Aún tomando en cuenta la falibilidad de estas prácticas – tal y como las he detallado en la primera sección de este capítulo –, su confiabilidad es mas alta que la altísima confiabilidad de siempre asumir que un número natural es primo. Uno debe ser muy cuidadoso, por lo tanto, cuando usa ejemplos matemáticos en experimentos mentales como los de Berker. Aun cuando nuestras prácticas matemáticas también son el producto de contingencias históricas, sus estándares epistémicos son poco sensibles al contexto. Es por ello que pensamos, correctamente, que para aceptar

una proposición matemática como verdadera debé exigirle a cualquier persona, en un contexto como el nuestro, nada menos que haberlo probado o aprendido de acuerdo a nuestros estándares de rigor epistémico actuales.³²

³². Además, el escoger la pregunta ¿es un número natural dado primo? como ejemplo Berker introduce tres nuevas fuentes de posible confusión: nuestra dificultad de pensar sobre dominios infinitos, el sesgo de representatividad y la asimetría entre el saber que un número es primo y saber que no es primo. Empecé por esta última. Efectivamente, es mas fácil saber que un número no es primo que saber que es primo, y conforme los números crecen, la diferencia se vuelve mas sustancial. Saber que el número 8 no es primo es muy sencillo, basta darse cuenta de que es par; saber que el número 7 es primo es un poco mas difícil, requiere darse cuenta de que no es par, ni divisible entre 3 y 5. Saber que el número 7918 no es primo es muy sencillo, basta darse cuenta de que es par; saber que el número 7919 es primo es mucho mas difícil. Saber que el número 393050634124102232869567034555427371542904834 no es primo es muy sencillo, saber que el número de Cullen 393050634124102232869567034555427371542904833 es primo ha sido muy difícil de establecer y es muy difícil de aprender. Esto significa que los recursos que necesitamos para determinar si un número natural es primo o no dependen *en gran parte* en de qué número natural esté hablando. Cuando se nos pide pensar, como hace Berker en su descripción del experimento mental, en los números naturales en general, inmediatamente se nos vienen a la mente número *extremadamente pequeños* dentro del inmenso dominio infinito de los números naturales. A Berker mismo le sucede cuando usa como ejemplos a los números 25, 604 y 7 en vez de números como 567034555427371542904834, 170141183460469231731687303715884105727 o 1317313771737997107113149157167179199311337347359389701709733739743751761769907937941953967971983991. En otras palabras, somos presas de lo que se conoce como el sesgo de disponibilidad, la tendencia que tenemos todos los humanos a considerar como normales el tipo de casos con los que nos hé encontrado mas comúnmente. Dado que es más probable que nos hayamos topado con números naturales como 7, 25 o 604, es decir, números excesivamente pequeños, que números de cientos o miles de dígitos, cuando pensamos en *un* número natural, el sesgo de disponibilidad nos hace pensar automáticamente en un número de éstos. Esta es otra razón por la cual, cuando se nos piden nuestras intuiciones sobre qué sería necesario para que la creencia de que

Cabe preguntarse, entonces, si el extraño Berker del experimento mental del Berker real es el tipo de agente al que debe aplicarse este estándar tan alto, o si mas bien es como el algoritmo de Roberta, para el cual es posible considerar aceptable y hasta virtuoso el simplemente asumir que cualquier número natural es primo. Como he mencionado, en nuestro contexto socio-histórico, estar justificado en tener una creencia matemática es haber participado de manera responsable en las prácticas epistémicas que constituyen a las matemáticas como una institución, es decir, haberlo aprendido o probado de manera competente. También he señalado que, dados los enormes recursos que la humanidad dedica y ha dedicado a esta empresa, sus estándares epistémicos – tanto los que demanda como los que alcanza – son altísimos, mas altos que los del proceso que persigue el extraño Berker del experimento de Berker. Lo que es importante determinar ahora es si es posible demandar de este extraño ser el participar de dichas prácticas.³³ Es decir si tiene disponibles los recursos necesarios para hacerlo. Es aquí donde cobra importancia la estipulación que introduce el Berker real de que éste es el único proceso cognitivo a través del cual el extraño Berker forma creencias sobre números primos. A través de esta estipulación el Berker real le cierra al extraño Berker la puerta que lo podría comunicar con la comunidad epistémica de los matemáticos, pero según contextualistas como yo, también lo exime de la responsabilidad de hacerlo. En otras palabras, dado que nuestras intuiciones sobre los estándares epistémicos apropiados para evaluar creencias matemáticas humanas dependen de consideraciones que no se aplican en el caso del extraño Berker, no nos dicen mucho sobre si la creencia de este extraño ser es conocimiento o no.

Es aquí donde el Berker real nos hace confundir nuestras intuiciones sobre los estándares aplicables a seres humanos como el Berker real, y aplicárselos a seres extraños como el Berker de su ejemplo. Es por ello que lo que nos muestra el experimento mental de Berker es que tenemos buenas razones (que tienen que ver con los enormes recursos dedicados a la matemática) para decir que quien (es decir, cualquier persona que) base su creencia de que 7 es primo en algo menos que un aprendizaje competente o una prueba matemáticas no está justificado en su creencia, lo cual es completamente consistente con el consecuencialismo, o por lo menos con las

³³. A decir verdad, hay casos genuinos en los cuales tiene sentido preguntarse de, por ejemplo, una persona con algunos tipos de discapacidades si efectivamente tiene la capacidad de participar en una institución social epistémica como la de las matemáticas.

tesis consecuencialistas básicas que adopto aquí, a saber, que una de las razones por las cuales preferimos ciertos medios sobre otros, es decir, uno de los criterios bajo los cuales evaluamos nuestros medios es por su capacidad a acercarnos a nuestros fines.

C. Suerte Semántica y Moral

El verdadero desorden
no consiste en tirar
sino en no recoger.
Alejandro Magallanes (2011, 97)

Los orígenes de este curso se encuentran en un artículo que publiqué en 2012 sobre las fuertes similitudes que encontraba en dos debates aparentemente muy distintos: el debate alrededor de la suerte moral en meta-ética y el debate alrededor del relativismo en la semántica de los modales epistémicos (es decir, expresiones aparentemente modales como “puede que” o “tal vez” cuando se usan para expresar lo que no es posible excluir a partir de lo que sabemos). Después de publicar dicho trabajo, la idea no desapareció de mi mente y poco a poco me fui dando cuenta que las similitudes eran más profundas y extensas de lo que había pensado originalmente. El propósito central de este curso ha sido precisamente extender los resultados de ese trabajo a otros campos y darles otras aplicaciones. Sin embargo, el ejemplo de los modales epistémicos me sigue pareciendo un muy buen ejemplo para ilustrar el debate entre deontologistas y consecuencialistas, además de dar pie a mi propia propuesta pluralista. La propuesta relativista de John MacFarlane que criticó en dicho artículo también me parece una excelente y sofisticada propuesta teórica que se toma en serio tanto al consecuencialismo como al deontologismo y criticarla me parece fundamental para fundamentar porqué, después de todo, no soy ni consecuencialista ni deontologista.

MacFarlane parte de la intuición básica de que, en muchas ocasiones, lo que nos interesa no está a nuestro alcance cuando lo necesitamos; esto nos obliga a tomar decisiones y hacer juicios en estados de indeterminación y con información limitada. Hasta aquí, MacFarlane sigue la ortodoxia consecuencialista. Sin

embargo, lo que distingue la posición de MacFarlane (y autores afines como March Richards) es que en ella, lo correcto incorrecto de estas decisiones y juicios dependerá de la información disponible, no tanto (o no solo) disponible al momento y para la persona que debe tomar la decisión o emitir el juicio, sino disponible al momento y a quién evalúa dicha decisión o juicio. Esto significa que la decisión será correcta o incorrecta dependiendo de la información que dispongamos para evaluarla, es decir, su corrección será sensible al contexto de evaluación. Según este tipo de relativismo, un acto podrá ser correcto desde una perspectiva, pero incorrecto desde otra; conforme conozcamos mas (o menos) sobre sus consecuencias, el mismo acto podrá pasar de correcto a incorrecto y vice versa. ¿Cuántas veces no nos he dado cuenta, conforme aprendí mas sobre una situación, que la decisión que tomamos no era la correcta? MacFarlane quiere hacer justicia de esta intuición y considera que el deontologismo no logra hacerlo. Para el deontologista, lo único importante es lo que uno sabe al momento de tener que tomar la decisión o hacer el juicio. Si mas adelante uno aprende mas sobre la situación, ése no debe cambiar la corrección o incorrección de nuestros actos pasados. En termino populares, para el deontologista “el hubiera no existe”. Si nos damos cuenta después de que las cosas hubieran sido mejor si hubiéramos actuado de manera distinta, esto no significa que lo que de hecho hicimos (o no) era incorrecto.

Consideraré los siguientes dos casos, aparentemente muy distintos:

- A. La madre de Sally entra en el dormitorio de su hija y la encuentra buscando algo debajo de la cama. "¿Qué pasa?", le pregunta, "¿qué estás buscando?" (1) "Mis gafas – responde Sally – puede que estén por aquí". Después de echar un largo vistazo debajo de la cama, Sally no encuentra ahí sus gafas; así que se dirige a buscar en otros lugares, no sin antes de decir, (2) "Uy, estaba equivocada."
- B. "El [segundo] ejemplo es el de un camionero; un camionero que va conduciendo su camión y que no se percata de estar atropellando a un niño que jugaba en la calle. Al conocer el suceso, el conductor sentirá una clase de lamentación por la muerte del niño que nadie más sentirá. Fue él mismo, al fin y al cabo, quien causó la muerte del niño. Ante este hecho, consideraríamos una reacción negativa, incluso inmoral, que el camionero no reaccionase con una profunda tristeza y con un fuerte pesar por la acción que ha cometido. Una reacción sana –según la terminología de Williams– consistiría en la sospecha por parte del agente de que había alguna cosa más que él podría haber hecho y que podría haber salvado la vida del niño: si

hubiese estado un poco más alerta, o conducido más despacio; si hubiese visto al niño jugando cerca del camino, etc. ...Esperamos, afirma, que la lamentación del agente se dé hasta en casos en los que no pensamos que el agente haya cometido ninguna falta. Si pensamos que el conductor no podía haber hecho nada más para impedir la muerte del niño, intentaré consolarlo; pero esto no reemplaza que tuviese una visión negativa del conductor si él no mostrase ningún tipo de lamentación, y se limitara a afirmar que ‘es un hecho terrible que haya sucedido, pero yo hice todo el que pude por evitarlo.’” (Rosell 2006, 151-2)

Pese a que ambos casos parecen ser radicalmente distintos, desde (2012) he sostenido que, sin embargo, ambos son sólo casos distintos del mismo (tipo de) problema, el problema de conciliar razones consecuencialistas y deontológicas.

Empecé analizando del primer ejemplo, el de Sally y por qué ha llamado la atención de filósofos del lenguaje como John MacFarlane. La clave es la extraña reacción de Sally tras no encontrar sus anteojos detrás de la cama. Su “Uy, estaba equivocada” nos suena muy normal, pero es difícil identificar exactamente en qué se había equivocado Sally. Sus palabras suenan a una disculpa, pero no es claro de qué error se está disculpando. Si vé lo que hizo y dijo – juzgar que sus anteojos podrían estar bajo la cama y buscarlos ahí – suena como un comportamiento correcto y racional, nada por lo que disculparse después.

Según MacFarlane, si lo que Sally afirma en (1) fue también de lo que se retractó en (2), debe haber sido algo que Sally ya sabía antes de mirar bajo la cama (algo como lo que (1) parece significar literalmente, es decir, que (3) sus anteojos podrían haber estado debajo de la cama), pero también algo de lo que ella podría retractarse correctamente después de buscarlas allí (presumiblemente, algo que después se dio cuenta era falso, como que (4) sus gafas, de hecho, estaban debajo de la cama o que, (5) tras buscar bajo la cama, llegaría a saber que sus gafas estaban debajo de la cama. Desafortunadamente, no existe ninguna proposición, en el sentido tradicional, que sea verdadera en el contexto de (1) (como (3)) pero falsa en el contexto de (2) (como (4) y (5)). Por lo tanto, concluyen los relativistas, necesitamos un nuevo tipo de proposición que sea sensible al contexto, y en particular, sensible lo que sabemos sobre la situación en el contexto desde el que la evaluamos (el cual, por supuesto, puede cambiar y no ser el mismo que el del hablante al momento de hacer la afirmación). Así, lo que Sally habría afirmado en (1) sería una proposición sensible a lo que Sally sabe o no sobre la ubicación de sus anteojos. Pero lo

que Sally sabe sobre la ubicación de sus anteojos cambia cuando mira debajo de la cama – pasa de no saber si están ahí sus anteojos a saber que no están ahí. Por lo tanto, el valor verdad de lo que dijo Sally – de la proposición que expresó al decir “Mis gafas puede que estén por aquí” cambiará de ser algo verdadero cuando Sally lo dijo, a falso cuando se dio cuenta de que las gafas no estaban ahí. ~~Así, dado que lo que Sally sabía antes de mirar bajo la cama era consistente con que sus anteojos estuvieran ahí, esta proposición era verdadera al momento de que Sally lo afirmó y Sally lo sabía, pero se volvió falsa una vez que cambió lo que Sally sabía sobre sus anteojos después de buscarlos debajo de la cama.~~ Desde esta nueva perspectiva, ya no era factible que sus anteojos estuvieran ahí, por lo que la misma proposición, desde este nuevo contexto, se volvió falsa.

Tradicionalmente, no pensamos que las proposiciones pueden cambiar de valor de verdad. una vez que algo es cierto, no puedo volverse falso. Si yo hoy digo que es domingo y mañana también digo que es domingo, hoy habré dicho algo verdadero y mañana habré dicho algo falso. Sin embargo, no hay ninguna proposición que haya dicho ayer que haya sido verdadera, pero al decirla hoy se haya convertido en falsa. No sería correcto describir lo que sucedió diciendo que ayer dije algo que era verdadero ayer, pero falso hoy. En su lugar, lo correcto es decir que ayer dije una proposición verdad y que, usando las mismas palabras, hoy dije una proposición distinta y, esta vez, falsa. Pero esto no es lo que quiere proponer MacFarlane. Lo que el propone es que lo que dijo Sally era primero verdadero y luego, la misma proposición, sin cambiar su sentido, se convirtió en falsa cuando Sally se dio cuenta de que sus anteojos no estaban bajo la cama.

Otra manera de modelar un relativismo de este tipo es como proponiendo que lo que pensábamos era una relación diádica – entre el objeto y la propiedad que juzgamos debe tener – es en realidad una relación triádica³⁴ – entre el objeto y la propiedad que juzgamos debe tener *y la perspectiva desde la cual lo evaluamos*. En este caso, la verdad, que solíamos pensar era una relación diádica entre la proposición y el mundo (la proposición es verdadera si el mundo es tal y como lo describa y falso si no) se convierte ahora entre una relación triádica entre la proposición, el mundo y la perspectiva desde cual la evaluamos (la proposición es verdadera si el mundo es tal y como lo describe desde la perspectiva en cuestión y falso si no).

³⁴. La propuesta mas famosa de este tipo es la teoría de las guisas de Nathan Salmon (1986)

Este tipo de propuestas, tradicionalmente conocidas como *Relativismos*, han sido consideradas absurdas en filosofía por lo menos desde los tiempos de Protágoras, y tratar de defender una propuesta de este tipo va a enfrentar siempre por lo menos dos tipos de problemas: problemas fundacionales y descriptivos. Los problemas fundacionales son determinar su una noción de proposición (y de verdad) de este tipo no es contradictoria, si puede integrarse al resto de nuestra semántica tradicional, y si es manejable de manera estable, etc. El debate descriptivo gira alrededor de la pregunta de qué tan bien el relativista da cuenta del tipo de fenómenos semánticos que putativamente la motivan. Por ejemplo, ¿qué tan bien explica el relativista lo que sucede en el caso de Sally? ¿Hay explicaciones alternativas? Y si las hay, ¿cómo se comparan sus pros y contras con las del relativismo?

El problema fundacional básico que enfrenta todo relativismo es que parece disolver la normatividad. Si lo correcto o incorrecto de un acto depende tanto de la perspectiva de la que se evalúa, ¿vas la pena preguntarse ¿de qué sirven ya nuestros juicios de qué es correcto o incorrecto? Tal parecería que si alguien ha hecho algo incorrecto, le bastaría solamente cambiar de perspectiva para que se vuelva algo correcto. Si mi perspectiva es diferente de la tuya, ¿cómo podríamos debatir de manera constructiva sobre lo que es correcto o incorrecto?

Gran parte del trabajo de MacFarlane ha tenido como objetivo explicar porqué estas críticas son injustas y se basan en una imagen equivocada de lo que es la normatividad. Para los críticos del relativismo, un relativismo como el de MacFarlane es inaceptable porque, si lo que afirmamos no es ni verdadero ni falso al momento en que lo afirmamos (como parece ser cada vez que decimos cosas como “Puede que mis gafas estén por aquí”, o en general, cosas sobre el futuro), no es posible comprometernos con su verdad y si no es posible comprometernos con su verdad, tampoco es posible afirmarlas, lo que desemboca en paradoja. Sin embargo, responde MacFarlane, es un prejuicio absurdo el pensar que no nos es posible comprometer con algo indeterminado. Por supuesto que es posible hacerlo y lo hice todo el tiempo. Cada vez que prometí hacer algo a futuro, por ejemplo, nos comprometí con algo aun indeterminado. A decir verdad, el hecho de que el futuro aun sea indeterminado es lo que da sentido a este tipo de promesas: nos comprometí precisamente porque aún es posible hacer algo. En este sentido, la moraleja de MacFarlane es que toda afirmación tiene algo de promesa, es

decir, su normatividad no se extingue al momento de que se hace la afirmación. En muchas ocasiones, afirmamos algo, es decir, nos comprometé con la verdad de algo precisamente porque aún es posible hacer algo al respecto. Aun cuando no podamos cambiar lo que ya hicimos, esto no significa que no quede nada que hacer. Aun es posible retractarnos, disculparnos, pedir perdón, reponer lo perdido, indemnizar, subsanar el daño, enmendar el error, defender nuestra decisión, etc.

Según (Barceló 2012), la reticencia que tenemos a aceptar *apriori* un relativismo como el de MacFarlane se debe a que nuestra concepción de la normatividad está guiada aún por principios deontológicos inaceptables y, en particular, por lo que en otros contextos se conoce como el “principio del control”, el principio según el cual no es posible ser responsables de aquello que no está bajo nuestro control, es decir, que la suerte no debe de tener consecuencias normativas. Esto significa que todos los factores relevantes para determinar si un acto es correcto o incorrecto, bueno o malo, deben estar determinados y disponible al agente al momento de actuar o tomar la decisión de actuar. En otras palabras, una vez que he actuado, *alea jacta est* y no hay vuelta atrás.

Contra este principio altamente intuitivo, MacFarlane no presenta otro principio opuesta y también intuitivo, sino que va a los casos particulares y muestra que, aunque el principio nos parece intuitivo, de hecho no actuamos en consonancia con él, sino que en nuestras prácticas normativas sí evaluamos negativamente actos en los cuales lo negativo no se debe a nada que estuviera bajo el control o acceso del agente al momento de actuar, sino a consecuencias que reconocí como imprevisibles. El ejemplo de Sally es uno, pero el de Williams es otro. Efectivamente consideraríamos una reacción negativa, incluso inmoral, que el camionero no reaccionase con una profunda tristeza y con un fuerte pesar por la acción que ha cometido aun en casos en los que no pensamos que el camionera haya cometida ninguna falta, sino que ha conducido con suficiente cuidado y responsabilidad. En otras palabras, pese a nuestras intuiciones deontológicas, no es posible imaginar ninguna circunstancia en la cual no valoraríamos negativamente a quien, tras causar la muerte de una o más personas, no se sintiera mal, disculpara o aceptará indemnizar a los afectados, sino que respondiera con un “ni modo, estas cosas pasan” o un “qué pena”. Aun más, es también fuerte nuestra intuición de que la manera en que debería de actuar el causante de dichas muertes es sustancialmente diferente de la manera en que debería comportarse cualquier persona, lo

mínimamente compasiva y empática, ante una desgracia de esta envergadura. En otras palabras no basta un acto de *condolencia* sino que se necesita un acto de *arrepentimiento*, un *mea culpa*.

Sin embargo, según el deontologismo, de hecho es cierto que “estas cosas pasan” y que hay circunstancias en las cuales el agente no tiene responsabilidad alguna sobre ciertas consecuencias imprevisibles de sus actos, no importa qué tan nefastas y negativas. Recordé que, según el deontologismo, estas consecuencias no podrían realmente atribuirse al agente, sino que sólo deberían considerarse efectos inconvenientes de actos irreprochables. Sin embargo, tenemos intuiciones fuertes de que nuestros actos pueden llegar a tener consecuencias tan negativas que no pueden considerarse meros efectos inconvenientes. Los casos que menciona Williams y Macfarlane son precisamente de este tipo.

En (2001) Andrew Latus ya había reconocido que el problema de la suerte moral puede presentarse como la paradójica conjunción de un principio general intuitivo y ciertos hechos morales particulares que lo contradicen:

- 1) La intuición de que la suerte no debe hacer diferencias morales (por ejemplo, que la suerte no debe afectar el valor moral de una persona ni aquello de lo que una persona es moralmente responsable).
- 2) el hecho (documentado empíricamente por Berg-Cross 1975, Cushman et.al. 2009, Cushman 2008, etc.) de que la suerte sí parece hacer diferencias morales (por ejemplo, culpamos al conductor desafortunado que causa un daño más que al afortunado conductor que pese a ser diligente llega a su destino sin mayores consecuencias).

La paradoja surge precisamente en los casos de suerte, ya sea buena o mala. En este tipo de casos, la intuición deontológica nos dice que la suerte no debe afectar nuestros juicios normativos. En los casos en los que el agente actúa con cuidado, por ejemplo, sigue siendo irreprochable su conducta aun cuando, por mala suerte, termine causando algún daño cualquier otro tipo de resultado negativo de sus actos. Sin embargo, es un hecho ampliamente documentado que cuando nos enfrentamos con este tipo de casos, nuestras intuiciones se vuelven consecuencialistas y consideramos que merece algún tipo de castigo – comúnmente una indemnización monetario a los afectados – el que causó daño, y que el cuidado con el que actúo tan sólo sirve, a lo más, para mitigar dicho castigo. En otras palabras, lo que tenemos aquí son actos para los cuales tenemos intuiciones fuertes

en sentidos opuesto e incompatibles. Para el deontologista, el agente actuó bien; para el consecuencialista, actuó mal. El trabajo del filósofo es tratar de disolver la paradoja.

Como en todos los casos en los cuales nos enfrentamos a posiciones en conflicto – una que dice que un objeto A es P (en este caso que un acto es bueno o mal) y otra que dice que A no es P – al filósofo se le abren varias opciones. Las dos primeras son las más obvias: defender una de las posiciones y criticar la otra. De hecho, en la literatura sobre normatividad, este tipo de respuestas suelen ser las más comunes. Para los consecuencialistas, la regla deontológica general de que la suerte no debe tener relevancia normativa es falsa y debe respetar nuestras prácticas normativas usuales; mientras que para el deontologista, dichas prácticas son erróneas, pues lo correcto es respetar la regla general de que sólo somos responsables de aquello que está bajo nuestro control y disposición.

Nótese que la misma paradoja que Latus identificó en el caso de la suerte moral se da también en casos como el de Sally, es decir, también ahí tenemos dos posiciones en conflicto:

- . 1) La intuición de que la suerte no debe hacer diferencias semánticas, por ejemplo, que la suerte no debe afectar de que es responsable una persona cuando hace una afirmación
- . 2) El hecho de que la suerte sí parece hacer diferencias semánticas (por ejemplo, pedimos disculpas por haber dicho algo falso, pese a que no teníamos manera de saberlo).

Al igual que en el caso de la suerte moral, en mi artículo de 2012, afirmé que las respuestas al problema que llamé “suerte semántica” suelen ser también de dos tipos. Algunos toman la perspectiva consecuencialista y afirman que la regla general está equivocada, que no hay nada malo en que la suerte afecte las condiciones de verdad de lo que decimos; mientras que otros defienden la posición deontológica de que la suerte nunca marca una diferencia en las condiciones de verdad de una aserción. Según William, por ejemplo, (i) el conductor del camión es (i) inocente, (ii) desafortunado, y (iii) justificadamente se siente mal después del hecho porque (iv) hizo algo horrible. Es (i) inocente porque el accidente no se debió a ninguna negligencia de su parte, sin embargo fué lo suficiente desafortunado (ii) para atropellar al niño. Además, (iii) está justificado en sentirse mal porque (iv) realmente hizo algo horrible: mató a un niño. Una posición análoga a la de Williams para el ejemplo de

MacFarlane sería decir que Sally es (i) inocente (al afirmar que las gafas pueden estar debajo del lecho no dijo nada malo), (ii) desafortunada (ya que las gafas, de hecho, no estaban allí), y (iii) justificada al retractarse *post facto* porque dijo algo que era, de hecho, (iv) falso.

Sin embargo, en 2015 me di cuenta de que el menú de opciones era en realidad mas amplio del que pensaba en (2012) y publiqué en mi blog una nueva clasificación de posibles posiciones que se pueden tomar, en general, en filosofía cuando nos enfrentamos a dos posiciones en tensión, cada una con buenos argumentos, intuiciones, etc. En otras palabras, tenemos buenas razones³⁵ para sostener que A es B y buenas razones para sostener que A no es B. Como ya he mencionado, existen las opciones obvias de defender unas razones y criticar las otras. Esto es lo que suelen hacer los deontologistas y consecuencialistas ya mencionados. Sin embargo, también existen por lo menos otras cuatro opciones para tratar de incorporar las *dos* posiciones en una simple teoría sin caer en contradicción (además, por supuesto, de la opción dialetheista de aceptar la contradicción, no como un problema que resolver, sino como una característica del fenómeno, como lo hace, por ejemplo, Graham Priest (1985)):

1. Gradualismo de propiedades
2. Dualismo de hechos/objetos
3. Dualismo de valores/propiedades
4. Relativismo

Analizaré detalladamente cada uno de ellos:

1. Gradualismo de propiedades

Una primera opción para conciliar estas dos posiciones es postular una gradación entre B y no-B, ya sea que se postule que la propiedad en cuestión B, en realidad no es una propiedad que se tiene o no se tiene, sino que es mas bien una propiedad gradual que se puede tener solo en cierto grado, o bien porque se piensa que, entre las

³⁵. De ahora en adelante hablaré de “razones” para hacer referencia a cualquier base sobre la cual se sostiene una posición filosófica, incluyendo argumentos, intuiciones, etc.

cosas que son *B* y las que no son *B* hay un tercer tipo de **casos intermedios** que no son ni *B* ni no-*B*. Por ejemplo, hay quienes han sostenido que la razón por la cual tenemos razones encontradas sobre sí, por ejemplo, el gato doméstico es un objeto no-artificial (ya que es una especie biológica de animal) o artificial (ya que fue creada artificialmente por el humano según sus propias preferencias) es porque la artificialidad es una cuestión de grado y entre lo artificial y lo no-artificial hay varios casos intermedios como los ya mencionados gatos domésticos, las bacterias hechas de bacterias, las uvas sin semilla, el fierro, etc. (Dennet 1990, Kroes 2012, Elder 2007, Sperber 2007, Grandy 2007; apud. Asse 2015, p. 39). La estrategia general es comparar los casos en los que tenemos razones encontradas con casos en los que no; así diríamos que los casos de razones encontradas son *menos-B* que aquellos en los que las intuiciones apuntan todas hacia *B*, y *más-B* que aquellos en que nuestras razones apuntan todas hacia no-*B*. En otras palabras, en el extremo de artificialidad tenemos casos como el ferrocarril o las flores de papel (para los cuales no tenemos razón alguna para pensar no son artificiales) y en el otro extremo, el de la no-artificialidad, tenemos objetos naturales como el sol o un árbol en medio del bosque (para los cuales no tenemos razón alguna para pensar que son artificiales); en medio, objetos como los gatos domésticos que son más artificiales que el sol, pero no tan artificiales como el ferrocarril. En el caso meta-ético, esto significaría que los actos que tienen consecuencias negativas no-intencionales son un caso intermedio entre los casos que tienen consecuencias negativas intencionales y los que no tienen consecuencias negativas. En el caso epistemológico, esto significa que los casos Gettier se encuentran en un punto intermedio de valor epistémico entre el conocimientos y la creencia verdadera no justificada o la creencia falsa justificada, etc.

En general, los gradualismos suelen ser inferiores a los pluralismo en valor explicativo, ya que en donde los gradualistas ven sólo una diferencia de grado, los pluralistas ven una distinción de tipo. En este sentido, los pluralismo suelen ser mas informativos que los gradualismos. Sin embargo, como bien sabemos, mayor información no implica siempre mayor capacidad explicativa. En muchos casos, la información extra no es relevante y, por lo tanto, el modelo mas simple es preferible. Lo mismo sucede con el debate entre gradualismo y pluralismos: para que el pluralismo sea genuinamente superior al gradualismo, las distinciones que introduce deben tener algún valor explicativo; si no es así, basta con notar que las diferencias pueden concebirse como de

grado. En tanto yo defiendo un pluralismo, espero mostrar que la distinción en la que se basa tiene valor explicativo.

2. Dualismo de objetos

La segunda opción es rechazar que efectivamente las dos posiciones se refieran al mismo objeto. Así, podríamos decir que lo que, según el principio deontológico es correcto **no es lo mismo que** aquello que, según nuestras razones consecuencialistas, es incorrecto. En el caso de Sally, por ejemplo, es posible fácil sostener que no hay tal cosa como lo que Sally dijo o hizo ya que siempre que decimos algo, en realidad, decimos y hacemos muchas cosas. Así pues, podríamos argumentar que aunque lo que afirmó era correcto – sus anteojos podrían haber estado bajo la cama – lo que dijo que podría ser cierto – que los anteojos estaban bajo la cama – era falso. En el caso del camionero de Williams, la manera en que el camionero manejaba era correcta, pero hizo algo horrible que fue atropellar a alguien. Dentro de la ética Aristotélica de la virtud no es raro encontrar por lo menos atisbos de un dualismo de este tipo, pues se piensa que no es lo mismo juzgar el carácter de una persona que sus actos (Doris 2002). Así pues, lo que sucede en el caso del camionero es que reconocí lo horrible que es atropellar a alguien, pero no es posible achacar dicho horror al camionero en tanto no revela ningún rasgo negativo de su carácter. En otras palabras, su **acto** es reprochable, mas no así su **carácter**. Algo análogo sucede en el caso de Sally: el acto tuvo claras consecuencias negativas – se perdió tiempo buscando infructuosamente debajo de la cama – pero esto no nos dice nada malo sobre Sally misma, es decir, sobre su carácter. Su comportamiento siguió siendo racional.

Este tipo de dualismo recupera puntos importantes, tanto del deontologismo como del consecuencialismo. En particular, permite dar cuenta de porqué información qué es relevante para la evaluación desde el punto de vista deontológico es diferente de la información qué es relevante para la evaluación desde el punto de vista consecuencialista. Como he señalado, para el deontologista lo importante normativamente es lo que sucede *previamente* al acto (es decir, cosas como las intenciones del agente, cuánta información relevante tenía disponible, qué tan bien preparado estaba, cómo había reaccionado en situaciones similares, etc.), mientras que para el consecuencialista lo importante son las consecuencias, es decir, lo que sucede *posteriormente* al acto (si se logran los objetivos del acto, si se afectan a otras personas, en especial, si se daña a alguien, etc.). El dualista da

cuenta de esto argumentando que, en realidad, se hablan de dos cosas distintas. Las primeras – las intenciones, disposiciones, la preparación, etc. – nos hablan del carácter del agente, su visión es mas amplia – mientras que las segundas nos dan información del acto en tanto suceso en el mundo, no como parte de la historia de comportamiento del agente. Ontológicamente son entidades de tipo muy distinto.

3. Dualismo de valores

Si uno acepta que es el mismo hecho o acto el que se juzga, ahora como aceptable, ahora como reprochable, aun tiene la salida de argumentar que en cada caso se juzga el acto bajo diferentes criterios, relativamente independientes. Esta ha sido la perspectiva que he adoptado en este libro, señalando que medios y metas se evalúan de manera distinta.³⁶ Desde la perspectiva de la meta, lo único que importa es si se alcanza o no; mientras que los medios se evalúan tomando en cuenta otros criterios, incluyendo su capacidad para acercarnos a nuestras metas, dados los recursos a nuestra disposición. La idea es rescatar las razones del deontologismo y las del consecuencialismo, pero arguyendo que cada una de ellas es verdadera respecto a un tipo de valor distinto.

Todo dualismo enfrenta dos retos importantes: en primer lugar, debe détrar que la distinción que establece no es ad-hoc, sino genuina, es decir, que no se está postulando nada más para resolver razones en conflicto sino que efectivamente esta describiendo una distinción genuina en el fenómeno. Esto significa que debe tener otras manifestaciones y explicar otros aspectos del fenómeno además de los involucrados en las razones en conflicto. En segundo lugar, además de fundamentar bien la distinción, tiene también que darle unidad, es decir, explicar cómo se relacionan ambos lados de la distinción. En este sentido, todo dualismo enfrenta dos retos simétricos y encontrados: por un lado, debe mostrar que cada lado de la distinción es autónomo y sustancialmente diferente del otro, pero también debe mostrar que ambos lados forman una unidad lo suficientemente homogénea como para haber generado la paradoja original.

³⁶. En esta sección presenté mi versión del dualismo de valores, aunque no es la única propuesta dualista en el mercado. Kvanvig (2005), por ejemplo, ha propuesto también un dualismo similar al mío.

El caso del dualismo de valores que he propuesto no es la excepción. Lo que he tratado de hacer a lo largo de este curso ha sido mostrar que la normatividad instrumental, propia de los medios (la normatividad de los deontologistas) y la normatividad trascendental, propia de los fines (la normatividad de los consecuencialistas) son efectivamente dos lados de un mismo fenómeno – la normatividad – pero lo suficientemente diferentes como para haber generado prácticas y teorías meta-normativas distintas (el deontologismo y el consecuencialismo respectivamente). La distinción que propongo, por lo tanto, no es ad-hoc, sino que da cuenta de muchas diferencias importantes que el debate entre deontologistas y consecuencialistas no ha hecho sino realzar: por un lado, tenemos una serie de prácticas normativas que he llamado instrumentales, sensibles a los recursos disponibles, que reconocen la falibilidad de nuestros medios y que apelan a información relevante, como las intenciones de los agentes, su carácter, sus capacidades y hábitos, su preparación, etc. Por el otro, tenemos otra serie de prácticas normativas trascendentes, es decir, indiferentes a las limitaciones de los agentes y a los recursos disponibles, donde lo que importan son las consecuencias, etc. Ambas prácticas son relativamente independientes, aunque están relacionadas y esta relación no es otra sino aquella en la que han insistido los confiabilistas, es decir, que si bien ambos valores son distintos, es razonable, en ausencia de mayor información, inferir que si se cumple uno, se cumple también el otro. En otras palabras, es una inferencia razonable, aunque derrotable, que si no sabes nada más sobre un medio más que suele producir consecuencias positivas, que suele llevarnos a alcanzar nuestras metas, inferir que es un medio virtuoso, confiable; y vice versa, si no sabes nada sobre un acto más que fue producido por un medio virtuoso, confiable, es sensato inferir que tendrá consecuencias positivas, que nos llevará a alcanzar nuestras metas. Sin embargo, conforme más sabes sobre el medio o sobre si, de hecho, se alcanzó o no el fin, estas inferencias pueden ser derrotadas, en tanto es posible que aún el medio más confiable produzca resultados negativos o que la meta haya sido alcanzada solamente por suerte. En otras palabras, existe la buena y la mala suerte.

En el caso de Sally, un dualismo como el que propongo nos permite decir que, si bien es cierto que aquello que Sally dijo de manera correcta antes de buscar sus anteojos debajo de la cama es lo mismo de lo que después se retractó, esto no lleva a ninguna contradicción porque la cuestión de qué está uno justificado en afirmar se resuelve apelando a criterios muy distintos a los criterios que determinan qué debe retractar. Es por

eso que es posible decir algo correcto (dada la información limitada que teníamos al momento de hacer la afirmación – en el caso de Sally, la poca información que tenía sobre la ubicación de sus anteojos) y después vernos en la obligación de retractarnos (una vez que tenemos mas información relevante). En general, es por eso que es posible hacer algo correcto, y luego vernos en la obligación de disculparnos por haberlo hecho (como sucede con el camionero de Williams) sin caer en ninguna contradicción. Siempre que evaluamos lo que alguien dijo, lo hacemos de acuerdo con al menos dos criterios: ¿está justificada la persona para afirmar lo que está afirmando? y ¿es lo que la persona está afirmando verdad? La mayoría de las veces, ambos criterios coinciden, por lo que no notamos. Casos como los de Sally son interesantes precisamente porque en ellos estos dos criterios divergen, lo que resulta en las razones contradictorias que nos parecían paradójicas, pero no lo son.

En otras palabras, lo que rechazo son tesis monistas cómo que existe **la** norma de la aserción o que los actos humanos son simplemente buenos o malos. En su lugar, propongo reconocer que hay varias maneras de evaluar los actos (y las afirmaciones, las creencias, etc.). Para cada valor trascendente v , existe un valor instrumental asociado j (de “justificación”) tal que si es apropiado juzgar una entidad a en cuanto a si es v o no, entonces también es apropiado juzgar si es j , donde j y v están relacionadas de manera tal que si algo es j pero no v , entonces nos enfrentamos a un caso de **mala** suerte, y si es v pero no por ser j , entonces nos enfrentamos a un caso de **buena** suerte.

Como había ya mencionado, hacer una afirmación como la de Sally es muy parecido a hacer una promesa o una apuesta en la medida en que no importa cuánta información tengamos, no es posible descartar todas las posibilidades de que lo que afirmamos sea falso. Siempre estamos asumiendo un riesgo, pero algunos riesgos son racionales (y por lo tanto estamos justificados en actuar sobre ellos), mientras que otros no lo son, y aún el menor de los riesgos puede resultar en consecuencias indeseables. En estos términos, Sally se retracta, no porque el riesgo que tomó no era racional, sino porque su apuesta no resultó ganadora.

Finalmente, vale la pena mencionar que llamar a mi posición un dualismo es confundente, pues sería mas preciso decir que mi posición es un pluralismo en tanto sostengo que hay por lo menos dos criterios para evaluar afirmaciones. Como he mencionado en el capítulo anterior, además de la verdad y la justificación,

también evaluamos las afirmaciones según el criterio del conocimiento, y gracias a Gettier, se sabe que este tercer criterio no puede reducirse a los dos criterios antes mencionados.

4. Relativismo

Según un relativista como MacFarlane o Egan, Sally tenía razón al decir que sus anteojos podrían estar debajo de la casa porque, dado lo que sabía antes de buscarlos ahí, esto era cierto, pero también tenía razón en pedir disculpas una vez que no los encontró ahí porque, desde su nueva situación epistémica, es decir, dada la nueva información que obtuvo al mirar debajo de la cama, lo que dijo era falso. La verdad o falsedad de lo que Sally dijo depende del conocimiento relevante desde el cual lo evaluamos, y este cambió para Sally del momento en que lo dijo al momento en que se retractó. Igualmente si lo que hizo el camionero es bueno o malo depende de la circunstancia desde la cual lo evaluamos, especialmente depende de cuanto y qué información tengamos disponible para realizar nuestro juicio moral. Desde la perspectiva de primera persona del camionero, él actuaba bien, estaba manejando con cuidado, pero nosotros no es posible verlo así porque nosotros estamos en una situación epistémicamente privilegiada en comparación, pues ahora sabemos que manejar con ese nivel de cuidado no era suficiente para evitar una desgracia. Ahora sabemos que el camionero debió haber manejado con mayor cuidado y desde nuestra situación sabemos que el camionero actuó mal. El acto pasó de estar bien en su perspectiva limitada, pero mal desde una perspectiva mas informada, aunque para entonces ya sea demasiado tarde, y no queda mas que disculparse.

Obsérvese que el relativismo y el dualismo son muy similares: ambos tratamos de romper el *impasse* entre deontologistas y consecuencialistas, tratando de encontrar lugar dentro de nuestra teoría tanto para la regla general de que la suerte no debe afectar cómo se juzgan nuestros actos, como para el hecho de que efectivamente hacemos cosas como disculparnos de consecuencias que estaban fuera de nuestro control. Pero lo hacemos de manera diferente. Para el dualista, la afirmación de Sally era correcta en un sentido (según los criterios instrumentales que guían nuestra acción) y equivocada en otro (según los criterios que determinan nuestros fines); mientras que el relativista cree que la afirmación de Sally era correcta e incorrecta bajo los mismos criterios (el

único criterio que rige tanto la aserción como la retracción, el cual tiene como miras los fines, pero es sensible a la información disponible en la situación desde la cual se hace la evaluación), pero en diferentes contextos: correcta dada la poca información relevante que Sally tenía cuando hizo la aserción, e incorrecta en la relativamente mejor informada situación en la que Sally se disculpó. Para el dualista, el camionero actuó de manera responsable cuando manejaba bajo los criterios responsables de cuidado que se demandan de alguien al manejar, pero actuó mal desde la perspectiva consecuencialista que toma en cuenta las consecuencias negativas de su acto. En contraste, para el relativista pensamos que el camionero actuó bien al manejar con lo que, de otra manera consideraríamos el suficiente cuidado; pero luego nos damos cuenta que, en realidad, no tuvo el cuidado necesario. En otras palabras, no estamos hablando de dos **tipos** distintos de evaluación, sino dos **situaciones** distintas de evaluación. Desde una perspectiva relativista, la suerte no es más que falta de información.

Para el relativista, el dualista exagera las diferencias entre la perspectiva que tiene el agente al actuar y la perspectiva que tenemos los que juzgamos su acto tras conocer sus consecuencias positivas y negativas; no se da cuenta de que estas son sólo dos posibles perspectivas desde las cuales podríamos evaluar un acto, pero en realidad hay muchas más. Consideré, por ejemplo, alguien que por alguna razón hubiera visto ya debajo de la cama antes que Sally y supiera que sus anteojos no están ahí y luego escuchará a Sally decirle a su mamá que pueden estar ahí. Si esa persona dijera “Sally piensa que sus anteojos están debajo de la cama, pero no es así” estaría en lo correcto y si dijera “Sally piensa que sus anteojos están debajo de la cama, y tiene razón” diría algo falso, y el relativista puede dar cuenta de ello de una manera que el deontologista no puede (MacFarlane 2014, 244). Para el deontologista, la única perspectiva relevante es la del agente al momento de actuar. Evaluar sus actos desde cualquier otra perspectiva sería injusto con ella. Sin embargo, lo hice. El relativista, en contraste, reconoce que, en muchos casos, la perspectiva más importante es la de quién hace la evaluación, pues reconoce que mientras lo que nos importe sea alcanzar nuestras metas, nuestra evaluación de los medios será siempre secundaria y condicionada por nuestra perspectiva. En otras palabras, la razón por la cual hice juicios con información limitada es porque no tenemos otra opción; si tuviéramos mayor información, sería absurdo no tomarla en cuenta. Es por ello que la persona que escucha a Sally desde una perspectiva mejor informada no

puede sino evaluar negativamente su aserción. No tiene mucho sentido ignorar lo que se sabe cuando es relevante.

Sin embargo, como MacFarlane mismo reconoce, no siempre es mas relevante la perspectiva del evaluador que la del agente, debido a que nuestras prácticas evaluativas suelen perseguir diferentes fines. Desafortunadamente, un monismo, aún uno tan flexible como el de MacFarlane no puede dar cuenta de esta multiplicidad. En este aspecto, el dualismo sí tiene una ventaja sobre el relativismo. El dualismo sí puede explicar porque, a veces tenemos la intuición de que, aun cuando sabemos mas que el agente, es posible *ponernos en su lugar* y entender qué lo que hizo era racional desde su perspectiva, mientras que en otras ocasiones, no es posible escapar nuestra perspectiva. Precisamente como lo que estamos evaluando son cosas distintas, por eso no es de extrañar que a veces es posible ver al mismo acto como correcto y otras como incorrecto. Aun más, dentro de una misma situación, es posible pasar de una a la otra y entender cuál es la perspectiva relevante en cada caso. Por ejemplo, como acertadamente señala Bradbury, no cualquier respuesta de contrición por parte del camionero de Williams sería adecuada. Es cierto que si sólo dijera algo como “¡Qué pena!” o “¡Qué tragedia!” es decir, si no asumiera una responsabilidad *personal* del acto, actuaría de manera reprochable, pero también es cierto que su reacción puede ser inapropiada también por ser *demasiado* personal, por ejemplo, si se culpa en exceso como si no hubiera atenuantes en su conducta y carácter. En ese caso, concluiríamos que el camionero ha confundido la respuesta adecuada a una consecuencia negativa relativamente imprevisible (disculparse, reconocer una responsabilidad débil) con la respuesta adecuada a un acto negligente (asumir una responsabilidad y culpa fuertes). Para MacFarlane ambas respuestas son idénticas: hacer lo posible ahora por lo que se hizo mal antes. pero la evidencia empírica de cómo de hecho actuamos y lo que esperamos de los otros en este tipo de situaciones parece ir en contra de lo que la teoría de MacFarlane predice y a favor de una propuesta dualista como la que propongo.

Por las mismas razones, el dualismo puede dar mejor cuenta de porqué, por ejemplo, es correcto responder a una disculpa como la de Sally con un “No te preocupes, no pasa nada” o “No podrías haberlo sabido”, etc. y como, al hacerlo, no señalamos que estuvo mal que Sally se disculpara. Responder así mantiene la

distinción entre los dos tipos de normatividad en los que he insistido en todo el curso: por un lado, reconoce lo apropiado de disculparse por las consecuencias negativas de nuestros actos, aun cuando sean imprevisibles, pero también reconoce lo virtuoso de actuar de manera cuidadosa aun cuando no se pueda excluir por completo la posibilidad de consecuencias negativas. Ni el deontologismo, ni el consecuencialismo, ni el relativismo pueden dar cuenta de este aspecto de nuestras prácticas normativas. Para el deontologista, efectivamente, dado que Sally no podría haber sabido que sus anteojos estaban debajo de la cama, no tendría porqué disculparse; para el relativista y el consecuencialista, lo importante era encontrar los anteojos y lo que Sally sabía o no es completamente irrelevante ahora que estamos en una mejor posición epistémica. Sólo el dualismo puede explicar que la disculpa de Sally es adecuada, por las razones que bien señalan consecuencialista y relativista, pero también es correcto señalar que Sally actuó bien, por las razones señaladas por los deontologistas. Así, el dualismo logra conciliar las posiciones deontológica y consecuencialista sin colapsarlas en una sola práctica.

D. ¿Realmente es diferente el relativismo al pluralismo?

Como he señalado unos párrafos atrás, una manera conocida de modelar el relativismo es como proponiendo que lo que pensábamos era una relación diádica – entre el objeto y la propiedad que juzgamos debe tener – es en realidad una relación triádica³⁷ – entre el objeto y la propiedad que juzgamos debe tener *y la perspectiva desde la cual lo evaluamos*. En otras palabras, si tenemos razones encontradas entre que un objeto *O* tiene una propiedad *P* y que no tiene dicha propiedad *P*, la manera en que el relativismo trata de resolver la tensión es introduciendo un

³⁷. La propuesta más famosa de este tipo es la teoría de las guisas de Nathan Salmon (1986)

tercer elemento contextual, C , tal que O sea P desde cierto C y no P desde otro C .³⁸ Por ejemplo, tenemos intuiciones muy fuertes e inconsistentes respecto a si el huitlacoche es delicioso o no; para muchos lo es, para otros, lo contrario. La mayoría de los filósofos adoptan un relativismo de lo delicioso según el cual, el huitlacoche es delicioso **para unos** y no delicioso **para otros**. Al introducir la perspectiva desde la cual juzgamos si algo es delicioso o no, resolví la tensión entre diferentes apreciaciones estéticas.

En contraste, el pluralista trata de resolver la misma tensión diciendo que donde parecía había un sólo objeto o una sola propiedad, en realidad hay dos o más. Para el pluralista, al igual que el relativista, la pregunta ¿es O P o no? está mal formulada; pero mientras que para el relativista la pregunta es incorrecta porque sólo se puede responderse en relación a un tercer elemento C , para el pluralista de objetos, la pregunta es incorrecta porque no es O lo que puede ser o no ser P sino una de otras dos o más entidades asociadas O_1, O_2, O_3 , etc. Lo mismo sucede en el caso del pluralismo de propiedades. Para este tipo de pluralista, la pregunta ¿es O P o no? está mal formulada porque P no es el tipo de propiedad que puede tener O , sino una de otras dos o más propiedades asociadas P_1, P_2, P_3 , etc.

Para ilustrar el pluralismo, pongamos un ejemplo de pluralismo de objetos exitoso (el caso de predicados es completamente análogo). Es tradicional pensar que propiedades cromáticas como “rojo” y “verde” son inconsistentes, es decir, que lo que es rojo, no puede ser también verde al mismo tiempo. Sin embargo, ¿qué responderíamos a alguien que presentase como contra-ejemplo a la sandía, la cual es tanto roja como verde? Lo más obvio sería responder que la sandía no es verde y roja, sino que la sandía *tiene partes* que son verdes y *partes* que son rojas (pero ninguna parte que sea verde y roja al mismo tiempo). En otras palabras, cuando decimos que la sandía es verde, en realidad queremos decir que la *cáscara* (o epidermis) de la sandía es verde; cuando decimos que la sandía es roja, en contraste, queremos decir que su *pulpa* (o endocarpio) es roja. En otras palabras, cuando parecía

³⁸. Este tercer elemento contextual puede ser cualquier cosa: el tiempo durante el cual el objeto puede ser o no ser P , la situación contextual particular desde la cual se juzga si es P , las convenciones sociales relevantes para determinar si es P , etc. Sin embargo, para ahorrar espacio usaré siempre la expresión “perspectiva”, a sabiendas que no todo relativismo es un perspectivismo.

que hablábamos de un sólo objeto – la sandía, en realidad hablábamos de dos – su cáscara y su pulpa. Una vez que distinguimos los dos objetos, la tensión se disuelve.

En este momento no se necesita ser demasiado suspicaz para darse cuenta de que la distinción entre pluralismo y relativismo parece haberse diluido, en tanto que parece que lo que ha hecho el pluralista no ha sido sino lo mismo que el relativista, es decir, introducir un tercer elemento respecto al cual el objeto puede tener o no la propiedad en cuestión. En este caso, ha introducido la noción de *parte* como un tercer elemento relativo a el cual lo sandía puede ser tanto verde como roja. Tal parece que lo que ha hecho el pluralista es sustituir la relación diádica entre objeto y color para convertirla en una relación triádica entre objeto, color y parte, lo cual es la definición misma de relativismo. Parece muy sencillo convertir cualquier relativismo en un pluralismo. Basta decir que cada vez que parece que adoptamos una perspectiva distinta, estamos hablando en realidad de un objeto distinto (el objeto desde la perspectiva) para convertir una teoría relativista en una teoría pluralista de objetos y que estamos hablando en realidad de una propiedad distinta (la propiedad desde la perspectiva). Y parece también muy sencillo convertir cualquier pluralismo en un relativismo. Basta decir que cada vez que hablamos de aparentes objetos o propiedades distintas, estamos en realidad hablando del mismo objeto y la misma propiedad, pero desde perspectivas distintas. En otras palabras, es posible fácilmente pasar de un relativismo que nos dice que un objeto O sólo puede tener una propiedad P relativo a una perspectiva C , a un pluralismo que nos diga que no es el objeto O mismo, sino *O-relativo-a-la-perspectiva-C* el que puede tener la propiedad P o a un pluralismo que nos diga que la propiedad que el objeto O puede o no poseer no es P *simpliciter*, sino *P-relativo-a-la-perspectiva-C*; e igualmente es posible fácilmente pasar de un pluralismo que nos diga que no es el objeto O mismo, sino una de otras dos o más entidades asociadas O_1, O_2, O_3 , etc. las que pueden tener la propiedad P a un relativismo que nos dice que un objeto O sólo puede tener una propiedad P relativo a aquella perspectiva que nos permite distinguir entre O_1, O_2, O_3 , etc. y de un pluralismo que nos diga que no es la propiedad P misma, sino una de otras dos o más propiedades asociadas P_1, P_2, P_3 , etc. las que puede tener el objeto O , a un relativismo que nos dice que un objeto O sólo puede tener una propiedad P relativo a aquella perspectiva que nos permite distinguir entre P_1, P_2, P_3 , etc. ¿Cuál entonces es la supuesta diferencia entre ser relativista y dualista?

La diferencia, por supuesto, depende de la naturaleza misma del tercer elemento en cuestión. Para que el relativismo sea una propuesta viable, este tercer elemento debe ser lo suficientemente autónomo como para jugar un papel del mismo nivel que los del objeto y la propiedad. Así por ejemplo, los filósofos aceptan un relativismo de lo delicioso en gran parte porque el tercer elemento relevante en su teoría – las personas – es algo que ya aceptamos que existe por razones independientes. En contraste, no pensamos en las partes como entidades independientes de aquello de lo que son parte. Por eso preferimos un pluralismo sobre el color en vez de un relativismo. Además, los objetos a los que apela el pluralismo – las partes del objeto – son entidades que también son aceptables dentro de nuestra ontología independientemente de cómo concibamos la relación entre colores y objetos. Estructuralmente, no hay gran diferencia entre las propuestas relativista y pluralista; la diferencia fundamental es en cómo dividen la realidad, es decir, en el tipo de entidades que forman su ontología, y para dar una explicación satisfactoria del fenómeno deben darnos buenas razones para aceptar que las entidades que postulan existen efectivamente y no son meras postulaciones ad-hoc para evitar caer en contradicciones o monismos.

El tipo de pluralismo que defiende en este curso es un pluralismo de propiedades. Sostiene que la pregunta por si un acto dado es bueno o malo, si una creencia es correcta o incorrecta, si un chiste es bueno o no, etc. están mal formadas, no porque exista un tercer elemento, una perspectiva de evaluación como la que postula MacFarlane, del cual depende si el acto es bueno o malo, la creencia correcta o incorrecta, el chiste bueno o no, etc. sino porque sostiene que detrás de evaluaciones tan generales como “bueno”, “malo”, “apto”, “correcto”, “incorrecto” etc. se esconden (por lo menos) tres propiedades de las que es posible estar hablando. Así pues, en casos de suerte (moral, epistémica, estética, etc.) lo que sucede es que tenemos razones en tensión sobre si el acto dado es bueno o malo, si la creencia es correcta o incorrecta, si el chiste es bueno o no, etc. y estas tensiones se pueden desenredar distinguiendo entre si lo que estamos evaluando es (1) si el acto alcanzó su fin trascendente, si (2) se actuó con responsabilidad y cuidado instrumental al actuar o si (3) se logró el fin trascendente porque se actuó con responsabilidad y cuidado. Estos tres tipos de propiedades son lo suficientemente diferentes e independientes como para no considerarse meramente tres diferentes maneras de apreciar que tan correcto o incorrecto es algo. A lo largo del curso he tratado de señalar como la diferencia entre estos tres tipos de

evaluaciones se manifiesta en diferentes aspectos de nuestras prácticas normativas, precisamente para distinguir mi propuesta de un relativismo como el de MacFarlane.

Para MacFarlane, en contraste, donde yo veo una distinción fuerte entre tres diferentes tipos de evaluaciones, él ve tres diferentes manifestaciones de la manera en que la perspectiva epistémica desde la cual se evalúa un acto afecta cómo lo hace. Así pues, el mismo acto puede ser correcto visto desde la perspectiva del que actúa, pero incorrecto desde la perspectiva de tercera persona de quién sólo ve sus consecuencias, y correcto otra vez desde la perspectiva de alguien más que conozca ambas cosas, etc. Desde un relativismo como el de MacFarlane, una propuesta como la mía exagera el hecho de que diferentes criterios y aspectos son relevantes para diferentes tipos de evaluaciones cuando lo que sucede es simplemente que evaluamos las cosas dependiendo de lo que sabemos de ellas. En otras palabras, para MacFarlane es muy importante poder dar cuenta del tipo de fenómenos, argumentos e intuiciones que da cuenta mi propuesta apelando a **un** sólo tercer elemento – lo que sabemos sobre el hecho desde el contexto en el que lo evaluamos – que no sea un mero constructo teórico de su propuesta, sino algo que es posible aceptar por lo menos *prima facie*. Y efectivamente, independientemente de lo que pensé de una teoría como la de MacFarlane es poco cuestionable pensar que desde diferentes perspectivas sabemos diferentes cosas sobre los hechos y que basamos nuestras evaluaciones en este conocimiento, independientemente de qué tan rico o pobre sea.

En contraste, mi propuesta sostiene que la situación del agente a la hora de actuar, la de la tercera persona que sabe si dicho acto alcanzó su objetivo y la de aquel quien sabe exactamente a qué se debe que el agente haya logrado su objetivo no son meramente tres diferentes perspectivas cada una con menos o más conocimiento del acto, sino que sus evaluaciones normativas son de tipo sustancialmente diferente. Quién señala que el chofer de Williams ha hecho algo horrible al atropellar a alguien está haciendo algo muy distinto de quién señala que se encontraba manejando de manera cuidadosa; no son dos personas haciendo lo mismo – evaluando moralmente el acto – desde diferentes perspectivas, como sostiene MacFarlane. Igualmente, quien dice de un chiste que es bueno, *porque* la gente no se rió y quien dice del mismo chiste que es malo *aunque* la gente se rió no están haciendo lo mismo – evaluando si el chiste fue bueno o no – desde diferentes perspectivas, como sostendría un relativismo, sino que están haciendo cosas distintas: uno está evaluando si el chiste alcanzó su objetivo de

hacer reír a la gente, y el otro está evaluando si el chiste era gracioso o no. Este ese punto central de diferencia entre un pluralismo como el mío y un relativismo como el de MacFarlane.

VI. Consecuencia

En todos los dominios de discurso y prácticas normativas la noción de consecuencia juega un papel fundamental. Sin embargo, capturar exactamente qué significa que algo sea consecuencia de otra cosa y, en particular, de una acción humana ha sido un problema gigantesco en la historia de la filosofía. En este capítulo defenderé también un pluralismo de consecuencias según el cual debe distinguir tres sentidos de consecuencia dentro de cada régimen normativo, correspondientes a los tres tipos de valores definidos a lo largo del curso. En otras palabras, cada vez que usamos un medio para alcanzar un fin, se establecen tres relaciones de consecuencia: una relación de consecuencia que transmite el valor trascendental de la meta, una relación de consecuencia que transmite el valor instrumental del medio, y una relación de consecuencia que transmite el logro de haber alcanzado la meta gracias al uso de un medio adecuado. En filosofía, las relaciones de consecuencia de primer tipo son harto conocidas y han sido muy estudiadas: consecuencia lógica (que transmite el fin trascendental de la verdad), deducción (que transmite el fin trascendental de tener una creencia verdadera), causalidad (que transmite el fin trascendental de que algo suceda), etc. Sin embargo, pese a que es muy sabido que estas mismas relaciones de consecuencia no transmiten los valores bajo los cuales evaluamos a los medios – justificación, responsabilidad, etc. – no contamos con teorías tan desarrolladas sobre qué tipo de relación de consecuencia *sí* transmitirían dichos valores. El objetivo de este capítulo no es desarrollar una teoría completa de la consecuencia para este tipo de valores pero sí trataré de, por lo menos, establecer una restricción necesaria para este tipo de relaciones, fundamentada en la respuesta a lo que he llamado la cuestión *fundacional* en el capítulo tercero. Para ello introduciré la noción de *derrotador* de una relación medio-fin. Recordé que uno de los criterios relevantes para juzgar un buen medio para cierto fin es que nos acerque a dicho fin y que esto se puede modelar como el estar apropiadamente preparado para, por lo menos, garantizar el éxito por lo menos en las situaciones racionalmente previsibles. Como señalé también en aquel capítulo, no toda situación posible es racionalmente previsible y eso es lo que hace que nuestros medios virtuosos sean falibles. Pues bien, si una de estas posibilidades no previsible se actualiza, decimos que la virtud del medio ha sido *derrotada* y a las posibilidades no previsible les llamaré “derrotadores”. Pues bien, la tesis principal que defenderé en este capítulo es que una noción adecuada de

consecuencia para medios debe respetar la derrotabilidad, es decir, que para que un medio M para un fin F sea consecuencia de otro medio M' para el mismo fin, es necesario que lo previsible para M sea también previsible para M' .

Esta restricción es bien conocida ya en el ámbito de la epistemología pues se usa para evitar lo que se Dretske llamo “pseudo-circularidades”. La idea fundamental es que la justificación ciertos estados doxásticos presuponen la ausencia de derrotadores y que, por lo tanto, no es apto usarlos para inferir la ausencia de esos mismos derrotadores. No seguir dicho principio nos haría caer en lo que es conocido actualmente como la paradoja del **conocimiento fácil** (Cohen 2002, Vogel 2000). Es bien sabido que el fenómeno del conocimiento fácil se encuentra detrás de muchas de las paradojas epistémicas mas conocidas asociadas a la falla de transmisión de justificación y conocimiento. Sin embargo, en este capítulo mostraré que el problema está mucho mas generalizado, ya que se da en todo ámbito normativo falible y que se encuentra detrás de paradojas aparentemente tan distintas como las del *sorites* y la del prefacio. Esto significa que una restricción similar a la propuesta por Dretske, Cohen, Vogel, etc. debe proponerse a toda relación de consecuencia que queramos transmita la adecuación de medios. Que no somos responsables de todo lo que hacemos implica es algo fácil de ver. Lo complicado es determinar qué, de lo que nuestros actos implican, sí somos responsables. No tengo una respuesta general a esta pregunta, pero sí puedo mostrar que no somos responsables de la ausencia o presencia de derrotadores de nuestra propia responsabilidad y que pensar de otra manera nos lleva constantemente a paradojas.

A. Derrotadores

Otra noción valiosa que es posible adoptar de la epistemología y adaptar a casos mas generales es la de derrotador. La idea básica es sencilla. En una sección anterior hacíamos la diferencia entre aquellas situaciones en la que una medida es exitosa y aquella en la que no. En un texto anterior (2017) había adoptado el termino de presuposición para hablar de las condiciones necesarias para que una medida sea exitosa. Desafortunadamente, el termino “presuposición” en lingüística y filosofía del lenguaje tiene ciertas connotaciones que ahora me parecen desafortunadas para habla del fenómeno que me interesa. En particular, la noción de “presuposición”

tiene la connotación de ser convencional y en ciertos debates se suele contrastar con cálculos de costo beneficio (por ejemplo, en Barceló y Stainton 2018). Afortunadamente, la noción que trataba de capturar con mi uso de “presuposición” puede capturarse de manera perfecta con otra noción dentro del cajón de herramientas de la filosofía actual de la cual es su opuesto perfecto: la de “derrotador”. Un derrotador es una condición suficiente para que una medida falle. De ahí que lo que antes solía llamar una “presuposición” sea lo opuesto a lo que tradicionalmente, y de ahora en adelante, llamaré un “derrotador”: una medida presupone que se van a satisfacer ciertas condiciones bajo las cuales dicha medida puede garantizar su éxito; si dichas condiciones no se satisfacen, es decir, si aparece un derrotador, desaparece dicha garantía – la relación entre la medida y su éxito garantizado es derrotada – y a menos que se aplique una nueva medida, en su lugar, nos queda la aleatoriedad y la suerte: si tenemos buena suerte, tendré éxito, si no, fallaré. Así es como se relacionan las nociones de suerte, presuposición, racionalidad, aleatoriedad, éxito, falla y derrota (de tal manera que cualquiera de ellas puede definirse a partir de las otras).

Añado la condición de “a menos que se aplique una nueva medida” porque es un fenómeno curioso, pero bien conocido que algo puede derrotar una medida para un fin *y al mismo tiempo* ser una nueva medida adecuada para ese mismo fin. En otras palabras, es posible que un derrotado derrote un medio que he usado para alcanzar un fin pero no nos deje desamparados y a merced de nuestra suerte, sino que al mismo tiempo nos provea de un nuevo medio (muchas veces, superior) para alcanzar ese mismo objetivo. Esto quedará claro una vez que ve algunos ejemplos.

¿Cómo se aplica esta noción a los ejemplos que he ya mencionado? Muy sencillo, si estamos preparados para una inundación que nos tenga aislados por una semana, el que la inundación no nos tenga aislados por más de una semana es la presuposición bajo la cual opera nuestra preparación, y si la inundación dura más, esto derrotaría la relación entre nuestra supervivencia y nuestra preparación. Una vez derrotada esa relación, nuestra supervivencia ya no depende de nuestra preparación. Igualmente, Mario Gómez-Torrente ha sostenido que el uso de términos vagos para referirse a grupos determinados de objetos presupone que el contexto de uso es uno en el que no hay series soríticas, lo que significa que las series soríticas son derrotadores de este tipo de uso de los

términos vagos. En otras palabras, poseer un concepto vago no debe entenderse en el sentido infalibilista de poseer la capacidad de, en cualquier situación, ser capaz de identificar la propiedad que aquellos objetos que caen sobre él tienen y los distingue de aquellos que no lo hacen. En su lugar, debe pensarse (por lo menos parcialmente) como la capacidad falible de, en situación previsible, ser capaces de identificar dicha propiedad o hacer dicha distinción. Por ejemplo, poseer el concepto de “montón de arena” no requiere tener la capacidad de, en cualquier situación ser capaz de distinguir un montón de arena de algo que no lo es, sino de ser capaz de hacerlo en situaciones normales, es decir, racionalmente previsible. En este sentido, el que no seamos capaces de hacer tal distinción en, por ejemplo, una serie sorítica no cuenta como evidencia de que no poseamos el concepto – o peor aún, de que el concepto sea deficiente en algún sentido – sino que sólo demuestra que somos falibles y que, por lo tanto, la herramientas que usamos para dar cuenta de la realidad también lo son.

Pensemos en los conceptos, las palabras, etc. entonces, como herramientas a la manera en que hablamos anteriormente, por ejemplo, de las medidas en contra de los ataques a civiles a través de camiones pesados o de medidas para prepararse en caso de riesgo de inundación. Así como es racional tomar medidas como prohibir la circulación de automóviles y colocar bloques de hormigón alrededor de las calles donde se esperan se junten multitudes, aunque sabemos que las medidas podrían evitar ciertos ataques, pero no todos, así también es racional usar conceptos o términos vagos a sabiendas que es posible usarlos para determinar una propiedad o extensión en ciertos contextos, pero no en todos. En otras palabras, sabemos que es posible usar la palabra “alto”, por ejemplo, para distinguir grupos de personas por su altura, aun cuando sabemos que hay situaciones en las que no nos servirá para dicho fin. Por ejemplo, si en la circunstancia hay una serie sorítica (es decir, entre cualquier par de personas A y B existe una serie de otras personas P_1, P_2, \dots, P_n , tales que $A=P_1, B=P_n$ y la diferencia de altura entre todo par secuencial de personas P_k y P_{k+1} (para todo k entre 1 y n) es tan pequeña que sería absurdo sostener que una es alta y la otra no) entonces nuestra capacidad para usar una descripción como “los altos” para referir a un grupo definido de personas en la situación queda derrotado (Gómez-Torrente 2010).

En epistemología, los derrotados muchas veces se manifiestan como información nueva que nos hace darnos cuenta de que las razones que teníamos para creer algo no eran tan buenas como creíamos. Un ejemplo

muy simple es enterarse de alguien que es un mentiroso. Al hacerlo, ya no es posible seguir creyendo lo que nos había dicho, aun cuando antes de enterarnos, estaba completamente justificado que creyéramos lo que nos había dicho. Igualmente, si tras dar un vistazo rápido al estacionamiento de nuestro trabajo reconocí el coche de una de nuestras colegas e inferimos de ello que vino a trabajar este día, nuestra creencia está justificada. Pero si luego nos enteramos de que otro de nuestros colegas se compró un coche del mismo color y modelo que el de nuestra otra colega, las razones que teníamos para creer que vino a trabajar dejan de ser suficientes para darnos justificación. En general, los ejemplos típicos de inferencias no monofónicas son todos casos en los que teníamos suficientes razones para sostener algo y luego, tener mas información hace que la información que teníamos antes deje de ser suficiente para la justificación.

Casos curiosos son aquellos en los que la nueva información derrota las razones que teníamos pero, al mismo tiempo, nos da *nuevas* razones para creer lo que ya creíamos. Supongamos que un amigo nos invita al cine. Quedamos de vernos en la entrada del cine, y cuando llegamos, nuestro amigo ya está esperándonos. Tras saludarnos, nos invita a que entré, dirigiéndonos hacia la entrada a la sala, y lejos de las taquillas. Esta información nos da buenas bases para creer que nuestro amigo ya tiene boletos y que, por lo tanto, ya es posible entrar a la sala. Nuestra creencia está justificada. Sin embargo, camino a la entrada de la sala vé un letrero que nos informa que la entrada es libre y los boletos no son necesarios. Nuestra antigua justificación para pensar que ya podíamos entrar a la sala queda derrotada, pero sorprendentemente no dejamos de estar justificados en creer lo ya que la información que recibimos del letrero no sólo derrota nuestra antigua justificación, sino que también sirve para justificar nuestra creencia. Derrota la justificación anterior, pero en cierto sentido, *se pone en su lugar* como nueva justificación.

Es importante remarcar que la derrota no necesariamente elimina aquello que nos justificaba, sino que hace que deje de justificarnos. Recordé que la justificación es altamente contextual, es decir, que el que cierta información, evidencia o razones sean suficientes para estar justificados en creer algo depende sustancialmente en nuestras circunstancias, si estas circunstancias cambian, lo que antes era suficiente puede ya no serlo. No es que nuestras razones previas desaparezcan, sino que dejan de proveernos de justificación.

B. Fallos de Clausura y Conocimiento Fácil

Si usted tuvo éxito, alguien en algún momento le ayudó. Tuvo un gran maestro en algún momento de su vida. Alguien ayudó a crear este sistema increíble ... que le permitió prosperar. Alguien invirtió en carreteras y puentes. Si tiene usted un negocio, no lo construyó usted [solo]. Alguien más lo hizo.

Barack Obama (2012)

Hasta este momento he estado pensando los fines los medios de una manera artificialmente atomizada, es decir, como si cada acto persiguiera **un** fin y para lograrlo usáramos **un** sólo medio, etc. Sin embargo, nuestra vida normativa no es así: establecí planes con fines a corto, largo y mediano plazo; cambiamos de opinión antes de alcanzar nuestros fines, etc. El objetivo de esta última sección del curso es arrojar luz sobre uno de los aspectos más simples de la compleja red de fines y medios en la que vivimos: el problema de cómo se transmite la justificación de los medios. Uno esperaría que si un medio es adecuado para alcanzar un estado dado E , también debería ser adecuado para alcanzar cualquier consecuencia de E . Sin embargo, toda la historia de la epistemología occidental está llena de contra-ejemplos a este simple principio. Pese a ser muchos y aparentemente diversos, al final, casi todos los famosos argumentos en contra del principio de clausura epistémica (como el de la mula disfrazada de zebra, el granero falso, etc.) muestran una misma estructura general de manera tal que todos terminan siendo instancias de un argumento general más abstracto conocido como el problema del “conocimiento fácil”.

Para entender por qué se le conoce así a este tipo de problemas vale la pena regresar a nuestra caracterización de la falibilidad de los medios. Habíamos dicho que un medio puede ser apto para perseguir un fin, aun si no garantiza que se alcanzara dicho fin. Esto se debe a que, para ahorrar recursos, es racional no tomar en cuenta ciertas posibilidades que, pese a estar efectivamente abiertas, son imprevisibles. Como también ya señalé, “imprevisible” aquí no debe entenderse solamente en el sentido en que no-conocible. En algunos casos, una situación calificará de imprevisible y, por lo tanto, soslayable, aun si es conocible o hasta conocida, siempre y cuando los recursos necesarios para prevenirla no están disponibles o no se justifica dedicárselos, porque su posibilidad es muy remota. Esto significa que por lo menos algunas posibilidades reales conocidas serán excluidas

del ámbito de lo previsible, porque son muy costosas. En otras palabras, es racional ignorar ciertas posibilidades porque no ignorarlas requeriría dedicarles demasiados de nuestros recursos y nos distraerá de lo que nos importa más. Las situaciones que dan pie a argumentos escépticos son casos claros de esto. Si tengo sed, por ejemplo, y tengo en mi mano lo que parece un vaso de agua, puede que me preocupe un poco saber si efectivamente es agua (no quiero terminar bebiendo algo tóxico, por ejemplo), pero me preocuparé menos por saber si efectivamente existe dicho vaso y no es sólo una simulación o el efecto de alguna estimación artificial de mi cerebro en una cubeta. Sé que si fuera un cerebro en una cuota no tendría buenas razones para creer que efectivamente hay un vaso de agua frente de mí, pero sé también que tratar de resolver ese problema requeriría un esfuerzo cognitivo superlativo, así que me concentro en la tarea mas razonable de averiguar si lo que contiene el vaso, asumiendo que exista, es agua o no y simplemente asumo que no soy un cerebro en una cubeta ni vivo en una simulación.

Sin embargo, el fenómeno es mucho más general. Si necesito calcular el área de una superficie, por ejemplo, uso las fórmulas que aprendí en la primaria, aunque sé que si la aritmética no fuera consistente dichas fórmulas no servirían de mucho; sin embargo, sé también que (1) demostrar la consistencia de la aritmética es una tarea titánica, (2) evidencia inductiva me dice que es muy poco probable que la aritmética sea inconsistente y de ahí concluyo que no vale la pena tratar de resolver el problema de la consistencia ahora y simplemente asumo que es consistente. Sé que mi reloj puede descomponerse en cualquier momento, pero cada vez que lo consulto no hago una verificación exhaustiva de su buen funcionamiento. Una vez mas, el esfuerzo requerido no vale la pena y, por eso, simplemente asumo que el reloj funciona y concentro mis esfuerzos en inferir qué hora es a partir de la información que me da.

Fuera del ámbito de lo epistémico, el fenómeno es igualmente general, pues comúnmente asumimos que no pasarán situaciones desafortunadas que sabemos pueden pasar, porque tomarlas en serio requeriría demasiado esfuerzo. Cuando organizo una fiesta para mis amigos, por ejemplo, trato de asegurarme que haya suficiente comida y bebida, que haya suficiente papel en el baño y que los invitados sepan como llegar y a qué hora, etc. Sé que todo eso contribuirá a que la fiesta sea divertida, pero sé también que existen muchas cosas que podrían arruinar mi fiesta y que puedo asumir no pasarán, no porque **sepa que no** pasarán, sino porque **no**

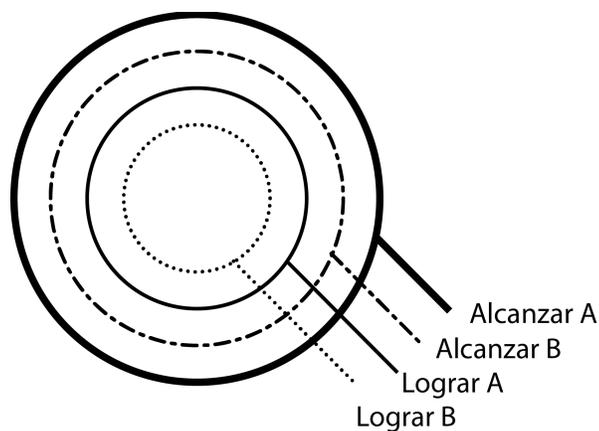
vale la pena preocuparse por ellas, y en algunas casos, la razón por la cual no vale la pena preocuparse por ellas es porque tomarlas en serio requeriría demasiado trabajo y otros costos que no es razonable dedicarles. Por ejemplo, sé que es posible que haya un desabasto de agua corriente y podría prepararme almacenando agua en recipientes, o instalando una cisterna, etc. Pero sé también que la posibilidad no es tan grande como para justificar todo ese trabajo y simplemente asumo que no se irá el agua.

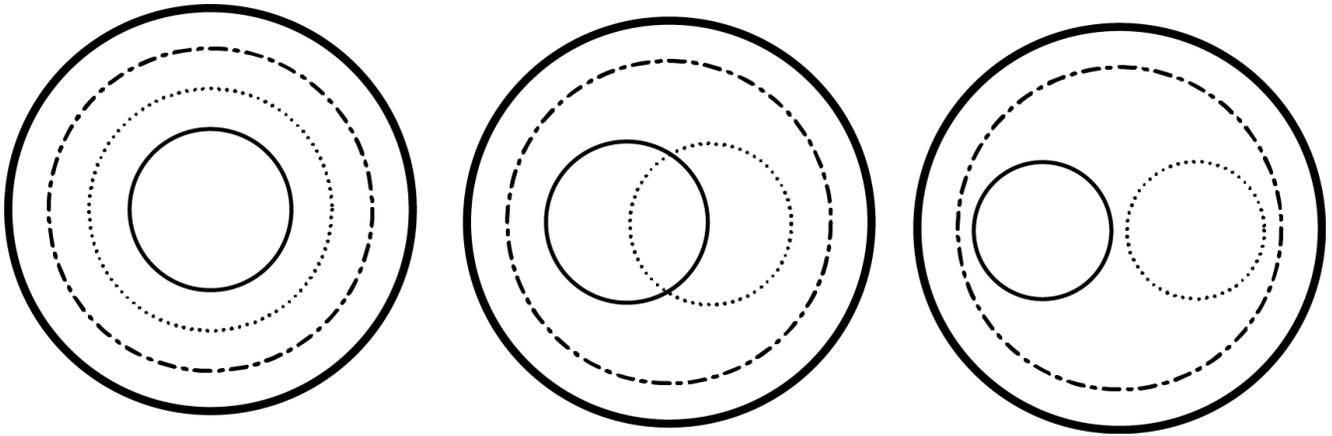
Esta distinción corresponde también a la distinción que filósofos como Stephen Yablo (2014) hacen entre las condiciones de verdad y *sobre qué trata* una proposición, aunque ideas similares se encuentran en el trabajo de, por ejemplo, Mitchell S. Green (1999). La idea general es que cuando aceptamos (o rechazamos, dudamos, creé, etc.) una proposición no nos comprometé con la verdad (o falsedad, o dudabilidad, etc.) de todas sus consecuencias lógicas, sino sólo con aquellas que tengan que ver con aquello de lo que la proposición trata. Por ejemplo, supongamos que Juan rechaza que Dolores sea chaparra porque piensa que su altura de 1.60 no se encuentra muy por debajo del promedio del país. Supongamos además, que Nuria considera que la expresión “chaparra” es ofensiva porque su contenido semántico contiene un estereotipo falso sobre las mujeres de cierta altura (Barceló, manuscrito). En consecuencia, Nuria rechaza que existan mujeres que encajen en el estereotipo y por lo tanto rechaza también que Dolores sea chaparra. Aunque ambos rechazan la misma afirmación, su rechazo tiene que ver con diferentes aspectos de la proposición; para cada uno de ellos la proposición trata de cosas distintas: para Juan, la proposición trata sobre Dolores y su altura en comparación con la de otras mujeres relevantes en su contexto; para Nuria, trata sobre la opresión de mujeres de cierta altura y, de Dolores en particular. Aunque es la misma proposición, hay un sentido en que sería correcto decir que lo que rechazan es algo distinto. Las consecuencias de cada rechazo, por lo tanto, también deben ser distintas. Juan puede inferir correctamente otras cosas sobre la altura de Dolores, por ejemplo, mientras que Nuria no. En otras palabras, las consecuencias de nuestros actos, estados epistémicos, etc. dependen de qué traten dichos actos.³⁹

³⁹. Decir simplemente que dependen de su contenido sería equívoco, precisamente porque tenemos **dos** maneras (por lo menos) de especificar el contenido de un estado cognitivo: a través de sus condiciones de verdad y a través de qué trata (su *aboutness*). La primera manera puede ser adecuada para nuestros fines consecuencialistas, pero no para nuestras evaluaciones deontológicas.

En otras palabras, saber que un derrotador está ausente suele ser una tarea difícil, por eso simplemente presupone que está ausente para poder pensar sobre otras cosas que nos son más urgentes. Desafortunadamente, al presuponer la ausencia de derrotadores, esta ausencia se vuelve una condición necesaria para el logro que obtené bajo su presuposición y, por lo tanto, se vuelve también consecuencia (por lo menos lógica) de dicho logro. Sin embargo, aunque sea condición necesaria para el logro de nuestros fines, no la tomamos en cuenta durante nuestra preparación y, por lo tanto, no es algo de lo que después podamos tomar crédito. Como célebremente expresó el senador estadounidense Barack Obama durante su campaña presidencial, una buena infraestructura de caminos, puentes, etc. es necesaria para el éxito de las empresas y, sin embargo, no es posible darles crédito a los empresarios de su existencia.

La falla de transmisión del logro a sus consecuencias lógicas es una consecuencia directa de la falibilidad de nuestro conocimiento. Como señalábamos en el capítulo anterior, esta falibilidad se puede diagramar usando dos curvas cerradas concéntricas: una representando todas las situaciones posibles en las que la meta se logra, y otra representando aquellas en las que la meta se alcanza. Dado que es posible alcanzar nuestras metas por suerte, la primera estará contenida en la otra. Como es bien sabido, la consecuencia lógica se puede representar también con dos curvas cerradas concéntricas, tales que si A es consecuencia lógica de B , toda situación que hace a B verdadera, también hace verdadera a A . Esto significa que la curva externa correspondiente a las situaciones que hacen a B verdadera se encuentra dentro de la curva externa correspondiente a las situaciones que hacen verdadera a A . Pero esto no nos dice nada sobre las curvas internas que representan las situaciones posibles en las que A y B no son sólo verdaderas, sino conocidas. En otras palabras, es posible tanto que una curva contenga a la otra, como vice versa; es mas, es posible que ninguna contenga a la otra, como es posible ver en los siguientes diagramas:





En este primer diagrama vé una posible distribución del espacio lógico en el cual A es consecuencia de B y el logro de A también es consecuencia de B ; desafortunadamente, hay muchas maneras en las cuales A puede ser consecuencia de B y aun así tengamos que el logro de B no se transmita a A :

Resumiendo, la ausencia de derrotadores es un claro ejemplo de falla de clausura normativa por consecuencia lógica (de los cuales la clausura epistémica es sólo un caso particular): la ausencia de derrotadores es condición necesaria de todo logro y, sin embargo, ella misma suele no ser un logro del agente. Para saber que la area de un triángulo es la mitad del producto de su base por su altura, es necesario que la geometría sea consistente, pero quien sabe que la area de un triángulo es la mitad del producto de su base por su altura no necesariamente sabe que la geometría es consistente. Para saber que tengo dos manos a partir de mi percepción de ellas frente a mí es necesario no ser un cerebro en una cubeta, pero quien sabe que tiene dos manos de esta manera no necesariamente sabe que es un cerebro en una cubeta. Para tener una empresa exitosa es necesario contar con una infraestructura pública apropiada, pero quién construye una empresa exitosa no necesariamente

construye también dicha infraestructura pública.⁴⁰ Todos estos son ejemplos de falla de clausura normativa y todos tienen la misma estructura.

Esquemáticamente, es posible reconstruir los argumentos de *falla de clausura / logro fácil* de la siguiente manera:

Premisa: Es posible lograr algo a través de un medio falible.

Por la premisa, existe un sujeto S (del cual asumimos es mínimamente competente), un fin P y un medio M tales que

1. S logró P. [Parte de la premisa]
2. El logro de S (de P) es falible. En otras palabras, es posible haber hecho lo mismo que S de hecho hizo para alcanzar P y aun así no obtener P [Parte de la premisa]
3. Sea D un derrotador del logro de S de P, es decir, una circunstancia posible en la que S ya no habría logrado P tras haber hecho lo mismo que de hecho hizo para alcanzar P.
4. De 2 y 3, para S, lograr P implica no-D, es decir, no-D es una consecuencia necesaria del logro de P.
5. Si Y es una consecuencia necesaria de X, entonces quien logra X logra también Y. [Principio de Clausura]
6. De 1, 4 y 5, S ha logrado que no D. [Consecuencia anti-intuitiva]

⁴⁰. Otra manera de entender el fenómeno es haciendo una distinción entre lo que una tarea **involucra normativamente** y lo que dicha tarea **implica lógica o metafísicamente**; lo que una tarea involucra normativamente es aquello que debé garantizar para lograr dicha tarea, así por ejemplo, organizar una fiesta involucra normativamente asegurarse de que haya suficiente bebida, pero no que no haya un desabasto de agua en el vecindario, aun cuando una buena fiesta requiere que haya agua corriente en la casa. Igualmente, creer justificadamente una proposición involucra normativamente asegurarse de que algunas de sus consecuencias lógicas sean verdaderas, pero no todas (incluso, como hé insistido a lo largo de todo el curso, no es necesario asegurarse de que la proposición misma sea verdadera). Construir una empresa exitosa involucra normativamente administrar de manera responsable los recursos, asegurarse un ligar robusto en el mercado, etc. pero no involucra construir carreteras para distribuir sus productos aun cuando sin ellas es imposible el éxito de la empresa.

Ergo, o bien ningún logro es falible o el logro no se transmite por consecuencia lógica. Obviamente, dado que me parece absurdo considerar que no haya logros falibles, me parece claro que la moraleja que debé sacar de la paradoja es que el logro no se transmite por consecuencia lógica. Ahora bien, si los logros no se transmiten a todas las consecuencia lógica, ¿a cuales sí se transmiten? En esta sesión espero haber dejado claro que no se transmiten a la **ausencia** de derrotadores del logro, aun cuando dicha ausencia es una consecuencia lógica de dicho logro. Sin embargo, creo que también es importante notar que por las mismas razones por las que el agente no puede adjudicarse como logro el que no haya habido derrotadores, tampoco puede adjudicarse su presencia (si no como logro, por lo menos como justificación). En otras palabras, así como lo logros no se transmiten a la ausencia de derrotadores, tampoco se transmiten a su **presencia** por las mismas razones, es decir, porque van mas allá de aquello de lo que el agente puede adjudicarse como logro. A fin de cuentas, si los derrotados se dan o no, es algo que no suele depender del agente y, por lo tanto, no puede adjudicársele como logro, aun cuando se siga lógicamente de lo que ha logrado. Ahora mostraré cómo no tomar esta restricción en cuenta nos lleva a caer en paradojas como las ya conocidas paradojas del prefacio, la lotería y el *sorites*.

D. Justificaciones Específicas y Generales

En una sesión anterior introducimos dos maneras para pensar sobre el contenido de una actitud proposicional una creencia o deseo, la aceptación o el rechazo de algo, etc.: una es atender sólomente a sus condiciones de verdad, es decir, a aquellas cosas de las que se sigue y que se siguen de ella de manera lógica; la otra es atender a lo que he llamado, siguiendo a Stephen Yablo, de lo que se **trata**. Pusimos el ejemplo de dos personas, Juan y Nuria, que rechazan ambos el que Dolores sea chaparra, pero por razones radicalmente distintas: Juan porque piensa que su altura de 1.60 no se encuentra muy por debajo del promedio del país, mientras que Nuria lo hace porque considera que la expresión “chaparra” contiene en su contenido semántico un estereotipo falso sobre las

mujeres de cierta altura, lo que la hace corresponder con un concepto vacío. Aunque ambos rechazan la misma proposición, para cada uno de ellos su rechazo trata de cosas distintas: para Juan, trata sobre Dolores y su altura; para Nuria, trata sobre la opresión de mujeres de cierta altura y, de Dolores en particular. Las consecuencias de cada rechazo, por lo tanto, también son distintas. Juan puede inferir correctamente otras cosas sobre la altura de Dolores, por ejemplo, mientras que Nuria no.

Como vimos claramente cuando estudiamos los casos de Gettier, algunas semana atrás, esta distinción aparece claramente también en el caso de las disyunciones explícitas y los enunciados existenciales. Dos personas distintas pueden creer de manera justificada que la disyunción 'P o Q' es verdadera por razones distintas – una porque cree que P es verdadera y otra porque cree que Q es verdadera. Aunque hay un sentido en el que ambas personas creen lo mismo, también hay un sentido importante en el que lo que creen es radicalmente distinto: en el primer caso, trata sobre P y en el segundo trata sobre Q; para una, aunque Q es parte importante de las condiciones de verdad de 'P o Q', no es parte de lo que ella trata; y viceversa para la otra. Aun más, es posible que haya una tercera persona que crea la misma disyunción sin tener suficiente información sobre si P o Q son verdaderas; es decir, sabe que por lo menos una de ellas es verdadera, pero no sabe cual. Para esta persona, 'P o Q' trata sobre algo radicalmente distinto. Mientras que la proposición que creen es la misma, su base para creerla es radicalmente distinta: en los primeros dos casos es información específica sobre P o sobre Q, mientras que en este tercer caso es información general sobre la disyunción.

Lo mismo sucede en el caso de proposiciones existenciales. Uno puede creerlas de manera justificada porque, o bien tiene información sobre un objeto específico que instancia la propiedad en cuestión o porque, en ausencia de tal información, sigue teniendo suficientes razones para creer que hay algo que satisface la propiedad aunque no pueda especificar qué objeto sea. Así, por ejemplo, yo puedo creer que hay cisnes negros o bien porque alguna vez me encontré con alguno o porque conozco la famosa historia de cómo la sociedad británica del siglo XVIII se asombró al ver por primera vez este tipo de aves, traídas por sus exploradores desde Australia. Otra vez, hay un sentido en el que en ambos casos creería lo mismo, también hay un sentido importante en el que lo que creería sería radicalmente distinto en tanto, en el primer caso, trataría sobre un cisne negro en específico y no en el segundo caso, sino más bien sobre la propiedad misma de ser un cisne negro, a saber, que no

está vacía. En general, proposiciones cuantificadas o disyuntivas pueden tener dos tipos de justificaciones: específicas o generales.⁴¹

Como vimos en el caso de “Dolores es chaparra” esta misma distinción existe también en el caso de las proposiciones singulares. Para Juan, “Dolores es chaparra” trata sobre el caso específico de Dolores, mientras que para Nuria trata sobre las personas a las que se les llama “chaparra” en general. Aun más sorprendente es que ocurre también en el caso de proposiciones universales. Por ejemplo, es posible creer una proposición universal como “mis sobrinos son traviesos” por dos tipos de razones: razones sobre los niños en específico que de hecho son mis sobrinos, o razones sobre mis sobrinos en general (por ejemplo, porque sabemos que todos los niños, en general, son traviesos). En otras palabras, nuestra justificación de proposiciones singulares o universales también puede ser específica o general.⁴²

Esto significa que es posible clasificar nuestras creencias (deseos, aceptaciones, etc.) según su contenido ya sea a partir de sus condiciones de verdad o a partir de lo que ellas tratan:

| CONTENIDO | Singular | Existencial | Universal |
|------------|---|---|--|
| Específico | Proposición singular cuya justificación está basada en razones específicas al objeto del que trata. | Proposición existencial cuya justificación está basada en razones específicas a un objeto que instancia la propiedad. | Proposición universal cuya justificación está basada en razones específicas a los objetos que instancian la propiedad. |

⁴¹. Esto no significa que la noción de *de qué trata* una actitud proposicional sólo importante sólo para distinguir entre justificaciones específicas y generales. Como ya mencionamos, dos personas que crean una proposición disyuntiva tipo ‘*P* o *Q*’, una porque cree que *P* es verdadera y la otra porque cree que *Q* es verdadera, creerán la misma proposición – con unas solas condiciones de verdad – pero sus creencias tratarán sobre algo diferente, y esta diferencia es importante para dar cuenta de casos como los de Gettier, aunque las justificaciones que puedan tener ambos sean las dos específicas, no generales. Sin embargo, para los objetivos de esta sección basta considerar sólo distinciones entre justificaciones específicas y generales.

⁴². Wittgentsein ya conocía esta distinción y ocupaba un lugar fundamental en su filosofía de las matemáticas (Barceló 2002) pues le permitía explicar la diferencia entre aplicar una regla e instanciar una generalización.

| CONTENIDO | Singular | Existencial | Universal |
|-----------|---|---|--|
| | Creo que <i>Manuel miente sobre sus síntomas</i> porque conozco sus síntomas reales y la he oído decir que son otros. | Creo que <i>hay un cisne negro</i> es verdadero porque ví uno. | Creo que <i>todos mis sobrinos son traviesos</i> porque los he visto comportarse así a ambos. |
| General | Proposición singular cuya justificación está basada en razones generales sobre la propiedad de la que se trata (de que es muy común o vacúa). | Proposición existencial cuya justificación está basada en razones generales sobre la propiedad de la que se trata (de que no es vacía). | Proposición universal cuya justificación está basada en razones generales sobre la propiedad de la que se trata (de que es vacúa). |
| | Creo que <i>Manuel miente sobre sus síntomas</i> porque sé que la gente, en general, miente sobre sus síntomas médicos. | Creo que <i>hay un cisne negro</i> es verdadero porque sé sobre su existencia en general, aunque no conozca ningún caso específico. | Creo que <i>todos mis sobrinos son traviesos</i> porque sé que los niños en general son traviesos. |

Según lo que vimos en la sesión anterior. Esta distinción es muy importante porque, si bien la primera manera de concebir su contenido determina sus consecuencias lógicas, la segunda es la que determina a cuales de dichas consecuencias se extiende nuestra justificación (y, consecuentemente, nuestro conocimiento). En otras palabras, para saber si de una creencia se sigue otra **en el sentido epistemológicamente** relevante de determinar qué inferencias transmiten justificación, no es suficiente determinar si las condiciones de verdad de cada creencia hacen a la inferencia lógicamente válida o no, sino que también es necesario determinar si de lo que tratan las premisas es también de lo que trata la conclusión. Si poné atención a la tabla anterior, esto significa que las inferencias que transfieren justificación sólo pueden darse entre creencias del mismo renglón, es decir, creencias con el mismo tipo de justificación: o específica o general. De premisas específicas (o una mezcla de premisas generales y específicas) sólo puede obtenerse una conclusión específica; y de premisas generales sólo puede obtenerse una conclusión general. Por ejemplo, si creo que *Manuel miente sobre sus síntomas* porque conozco sus síntomas reales y la he oído decir que son otros, y de ahí infiero la conclusión existencial de que *hay quien miente sobre sus síntomas*, ésta estara basada también en conocimiento específico sobre un objeto que instancia la

propiedad en cuestión; igualmente, si de dicha creencia, en conjunción con otras del mismo tipo, concluyo que *todos mienten sobre sus síntomas*, ésta conclusión heredera su carácter específico. Lo mismo sucede si mi justificación es general en vez de particular. Si creo que *Manuel miente sobre sus síntomas* porque sé que la gente, en general, miente sobre sus síntomas médicos, y de ahí infero la conclusión existencial de que *hay quien miente sobre sus síntomas*, ésta estara basada también en conocimiento general sobre la propiedad relevante, es decir, sobre el hecho general de que la gente miente sobre sus síntomas médicos.

Esta distinción es importante porque, como mostraré a continuación, está detrás de algunas famosas paradojas como son la del prefacio, la de la lotería y la de sorites.

E. Cómo pensar sobre otra cosa

El objetivo de esta última sesión es ofrecer una manera de resolver ciertas paradojas clásicas, aplicando algunas de las ideas que he desarrollado a lo largo de este libro. Mi solución está basada en la distinción que introduje en una sección pasada entre de lo que las proposiciones tratan y sus condiciones de verdad. La idea fundamental es que los logros no se transfieren a todas las consecuencias de lo que se logra, sino sólo a aquellas consecuencias con las que tiene que ver el logro. Esto significa, en particular, que dado que la presencia o ausencia de derrotadores no es parte de lo que trata un logro, no es algo de lo que pueda cobrar crédito el agente, aun cuando sea condición necesaria de lo que efectivamente logró. En particular, veré que en cada una de estas paradojas no está lo suficientemente descrita, pues sólo nos da las premisas y las conclusiones pero no nos dice de qué tratan. Peor aún y como argumentaré, no hay manera de fijar de qué tratan las premisas de manera tal que sean al mismo tiempo aceptables y se siga de ellas la presunta conclusión contradictoria que da pie a la paradoja.

Antes de entrar de lleno en mi propuesta, vale la pena decir un par de cosas sobre lo que quiero decir por ‘resolver’ una paradoja como ésta, es decir, una en la que, por lo menos *prima facie*, una conclusión contradictoria se sigue de premisas aceptables. Por principio de cuenta, hay tres maneras de resolver la paradoja: mostrando o bien (i) que la conclusión (S) no es realmente contradictoria, o (ii) que (S) no se sigue de las premisas (P), (Q) y (R), o que (iii) alguna de dichas premisas no es realmente aceptable. En cada caso, además, es necesario explicar o bien (i) porqué la conclusión, pese a no ser contradictoria, nos *parece* serlo, o (ii) porqué, aunque no se sigue,

parece seguirse de las premisas, o bien (iii) porqué, aunque alguna de las premisas no es aceptable, todas *nos parecen* aceptables. Soluciones de cada uno de estos tipos existen en la literatura. Mi solución es del segundo tipo, trataré de mostrar que, en todos los casos, la conclusión no se sigue de las premisas, aunque parezca hacerlo. No es de sorprender que, dado mi dualismo, mi diagnóstico será que la conclusión parece seguirse de las premisas, aunque de hecho no lo haga, porque se sigue en un sentido, no en otro. Se sigue en el sentido de que la conclusión es consecuencia lógica de las premisas, lo que corresponde al régimen de la verdad, es decir, el régimen de de los fines; pero no se sigue en el sentido de que la justificación de las premisas no se extiende a la conclusión, lo que corresponde al régimen de la justificación, es decir, el régimen de los medios.

Empezaré por una paradoja muy simple: la famosa paradoja del prefacio, la cual es muy sencilla de formular. Supongamos que un autor –llamémosle “Miguel”– escribe un libro de no-ficción. Miguel es un investigador concienzudo y ha hecho una labor formidable investigando su tema, pero también es sensato y sabe que pese a que lo que ha escrito es el resultado de una investigación seria y profunda, lo más probable sea que haya cometido por lo menos algún error y, en consecuencia, en su prefacio reconoce que por lo menos algo de lo que sostiene en el libro muy probablemente sea falso. No hay nada extraño, ni insensato en esto. Sin embargo, la actitud de Miguel frente a su propio libro parece ser paradójica. Por un lado, tiene la creencia existencial de que por lo menos una de las oraciones que sostiene en su libro es falsa, pero si le preguntáramos, de cualquier oración en su libro si es verdadera, sin duda el autor afirmaría su verdad sin chistar. En otras palabras, sean P_2, P_3, \dots, P_n las oraciones contenidas en su libro, el autor aceptaría que " P_1 es verdadera", " P_2 es verdadera"... o " P_n verdadera". Esto parece significar que piensa que todas y cada una de las afirmaciones de su libro son verdaderas, lo cual es inconsistente con su propia afirmación en el prefacio de su obra de que por lo menos alguna de dichas afirmaciones es falsa.

Esquemáticamente, el argumento tiene la siguiente forma:

1. La primera afirmación del libro es verdadera. [Premisa]
2. La segunda afirmación del libro es verdadera. [Premisa]
3. La tercera afirmación del libro es verdadera. [Premisa]

...

- N. La n afirmación del libro es verdadera. [Premisa]
- N+1. Hay n afirmaciones en el libro. [Premisa]
- N+2. Por lo menos una de las afirmaciones del libro es falsa. [Premisa]
- N+3. De 1 a N+1 se sigue que todas las afirmaciones del libro son verdaderas.
- N+4. Pero N+2 y N+3 se contradicen entre sí.

La paradoja del prefacio nos enfrenta con un dilema vergonzoso ambos de cuyos cuernos son inaceptables. O bien el autor de ningún libro puede realmente reconocer que debe haber por lo menos un error en su libro (y decirlo sólo es falsa molestia) o bien todo autor introduce en sus libros algo que no cree que es verdadero. Sin embargo, nos parece muy claro que es posible que uno puede racionalmente llenar un libro de afirmaciones que considera verdaderas y, sin embargo, reconocer que por lo menos una de ellas puede ser falsa, sin caer en una contradicción. ¿Qué sucede en esta situación entonces?

Mi propuesta es ver en la paradoja del prefacio una tensión entre un juicio existencial positivo y una serie de juicios singulares negativos que parecen excluir la instanciación del existencial: Tenemos un juicio existencial – hay por lo menos una afirmación falsa en este libro – en aparente tensión con una serie de juicios singulares – esta afirmación es verdadera, esta también, y esta también... – que parecen excluir la instanciación del existencial. Ahora bien, como he argumentado en la sección anterior, el que las premisas y la conclusión sea existenciales o singulares no nos dice nada sobre el aspecto de su contenido que realmente es relevante para determinar con cuales de sus consecuencias estamos racionalmente comprometidos, a saber, no nos dice nada sobre el tipo de justificación que tienen. Como señalamos en la sección anterior, un juicio singular como “La décimo tercera afirmación del libro es verdadera” puede tener diferentes tipos de justificación. Podemos aceptarla porque tenemos información específica sobre esta afirmación en particular, por ejemplo, si consideramos lo que dice dicha afirmación y evaluamos nuestras razones para sostenerla, o podremos aceptarla porque tenemos información acerca del libro que la contiene que nos hace confiar en lo que ahí dice. Tomemos un ejemplo real y supongamos que el libro fuera *Jusnaturalismo procedimental, debido proceso penal y epistemología jurídica* de Edgar Aguilera García (2017). Lo abro en una página al azar –la 67– y leo ahí que, *grosso modo*, es lo mismo

decir que el estar de derecho es o posee un valor moral y decir que contribuye a la realización o materialización de valores morales. Dado que mi conocimiento del derecho y su filosofía es muy limitado, pero conozco el trabajo del autor y he sido testigo de primera mano de sus habilidades filosóficas y su seriedad como investigador, adopto esta creencia a partir de su testimonio. La justificación que así obtengo, es muy distinta de la que podría tener otro abogado, jurista o filósofo del derecho (o Edgar Aguilera mismo) que sí conozca sobre el valor moral del estado de derecho. Mientras que yo tengo una justificación indirecta y basada en una confianza general en el autor y lo que ha escrito, no tengo una justificación específica a lo que dice dicha afirmación. Un jurista o filósofo como el autor, en contraste, sí tiene disponible esta justificación específica. Así pues, podemos terminar en estados epistémicos similares –ambos tendremos la misma creencia justificada– pero nuestra justificación sería tan radicalmente distinta, que no podríamos extender nuestra justificación a las mismas consecuencias de aquello que creemos. Por ejemplo, el jurista con justificación específica sobre el contenido de la afirmación puede usar dicha justificación para concluir, al leer el libro, que el autor es confiable (por lo menos en ese punto). Sin embargo, si yo hiciera esa misma inferencia, estaría incurriendo en una petición de principio falaz. Esto se debe a que si el autor no fuera confiable en este punto, mi justificación quedaría derrotada; por lo tanto, no puedo usar esta información para concluir que no se da dicho derrotador.

Regresando a la paradoja, cada vez que le preguntamos al autor **sobre** una de sus afirmaciones, la pregunta trata sobre dicha pregunta, no sobre el libro. La proposición “esta enunciado es verdadero” tiene entre sus condiciones (suficientes) de verdad la de que todos los enunciados del libro son verdaderos; pero es una posibilidad muy lejana desde la perspectiva del autor, en tanto cree que no todo lo que escribió es correcto. En su lugar, su justificación se basa en su conocimiento especializado en el tema del que escribió. Para él, de eso se trata cada premisa de la paradoja, no del libro.

Ahora bien, si tengo razón, y la manera en que Miguel acepta las premisas de la paradoja es bajo una justificación específica, entonces la conclusión $N+3$ que se sigue de ellas no puede sino ser también específica. Es decir, debe ser una conclusión que trate no sobre el libro, sino sobre el contenido de las afirmaciones que contiene. En otras palabras, la $N+3$ que se sigue de las premisas que Miguel acepta es una proposición acerca de aquello de lo que trata su libro, y al decir que Miguel sabe que es verdadera, lo que estamos reportando es que

Miguel sabe sobre aquello de lo que su libro trata. En cambio, la afirmación $\mathcal{N}+2$, la afirmación general de que no todas las afirmaciones que contiene el libro son verdaderas, no trata de lo mismo de lo que trata $\mathcal{N}+3$, a saber, de aquello de lo que trata el libro de Miguel, sino que trata de Miguel mismo y de su falibilidad. Por lo tanto, ambas proposiciones no entran en contradicción directa. Esto significa que aunque la conclusión contradictoria $\mathcal{N}+4$ se sigue lógicamente de $\mathcal{N}+3$ y $\mathcal{N}+2$, la justificación que Miguel tiene para $\mathcal{N}+3$ y $\mathcal{N}+2$ no se transmite a través de dicha relación de consecuencia, es decir, no se transmite a $\mathcal{N}+4$. Esto se puede demostrar a través del siguiente argumento:

Para que Miguel estuviera justificado en creer $\mathcal{N}+4$ tras inferirla de manera competente de $\mathcal{N}+3$ y $\mathcal{N}+2$, su justificación debería de ser o bien específica o general. Sin embargo, Miguel no tiene una justificación específica para $\mathcal{N}+3$ y $\mathcal{N}+2$, ya que la única justificación que tiene para $\mathcal{N}+2$ es general, es decir, su única justificación para creer que por lo menos una de las afirmaciones del libro es falsa está basada en el conocimiento de su propia falibilidad; de manera simétrica, Miguel tampoco tiene una justificación general para $\mathcal{N}+3$ y $\mathcal{N}+2$, ya que la única justificación que tiene para $\mathcal{N}+3$ es específica, es decir, su única justificación para creer de todo lo que afirma en el libro que es verdadero es su profundo y extenso conocimiento del tema del que trata su libro. En consecuencia, no hay nada de lo que podría tratar $\mathcal{N}+4$ que se siguiera de las premisas aceptables para Miguel. Así pues, Miguel no está comprometido con la conclusión contradictoria, aunque se siga lógicamente de cosas que cree de manera justificada. Así se disuelve la paradoja del prefacio.

F. Cómo resolver la Paradoja de la Lotería

Una paradoja de estructura muy similar es la paradoja de la lotería, según la cual hay una tensión entre creer que un boleto de la lotería ganará, y rechazar cualquier afirmación de que algún boleto particular será el ganador. Así es como la presenta Mauricio Zululaga (2005):

Supongamos que hemos comprado un billete de lotería, alguno de los cuales deberá ganar. Supongamos además que hay 100 billetes de lotería. La probabilidad de que uno de los billetes gane es muy baja: 0,01. Supongamos que usted ha comprado un billete de lotería. De acuerdo con la baja probabilidad de que su billete gane, usted está justificado para afirmar que creer que perderá —suponiendo que uno está justificado para creer que algo será el caso, si la probabilidad

de que ocurra es mayor a 0,5—. Pero bajo estas condiciones usted también está justificado para creer que cada uno de los billetes perderá. Así es racional creer que ninguno de los billetes ganará, porque para cada uno de ellos usted ha podido establecer que la probabilidad de que gane es menor a 0,5. Pero, *ex hypothesi*, ha de haber un billete de lotería que gane, así que usted cree algo contradictorio. Usted está justificado para creer que ningún billete ganará y, al mismo tiempo, cree que uno lo hará.

Esquemáticamente, el argumento tiene la siguiente forma:

1. El primer boleto de lotería no va a ganar. [Premisa]
2. El segundo boleto de lotería no va a ganar. [Premisa]
3. El tercer boleto de lotería no va a ganar. [Premisa]
- ...
- N. El n boleto de lotería no va a ganar. [Premisa]
- N+1. Hay n boletos de lotería. [Premisa]
- N+2. Un boleto de lotería va a ganar. [Premisa]
- N+3. De 1 a N+1 se sigue que ningún boleto de lotería va a ganar.
- N+4. Pero N+2 y N+3 se contradicen entre sí.

Esta paradoja está especialmente dirigida hacia el falibilista, para quien toda justificación es falible y por lo tanto debe reconocer que existe (por lo menos) un número para el cual, si sabemos que una lotería tiene ese número de boletos, entonces automáticamente estamos justificados a creer, para cualquier boleto de lotería, que dicho boleto **no** será el ganador, aun en ausencia de cualquier otra información sobre qué boleto ganará (es decir, que no es posible justificar una proposición particular de este tipo en base de lo que he llamado aquí una justificación general). Como he mencionado, es posible rechazar esta hipótesis y sostener que, en el caso de hechos de este tipo, conocer la probabilidad de un hecho, no importa qué tan alta, no es suficiente para conferir justificación a la creencia, pero me parece que no es necesario, pues es posible aplicar el mismo principio que aplicamos a la paradoja anterior para resolver también ésta.

Una vez más, el problema es que la paradoja pretende contraponer dos proposiciones que parecen contradecirse (en tanto sus condiciones de verdad son contradictorias) pero que, en realidad, tratan de cosas

distintas. Esto se debe a que las dos paradojas que he mencionado hasta ahora tienen la misma estructura. Empiezan señalando que es posible (y hasta normal) que una gente racional acepte tanto una proposición existencial como una serie exhaustiva de proposiciones singulares sobre el dominio del existencial de dicha proposición, pese a que estas proposiciones singulares son lógicamente inconsistentes con dicho existencial. Mi solución muestra que, efectivamente, hay una razón por la cual la justificación no se transmite de las premisas a la conclusión de manera que el agente esté comprometido a aceptar una contradicción. En otras palabras, de (1) y (2) no se sigue que S sea irracional:

1. S está justificado en creer que hay (por lo menos) un X que es F .
2. Para todo x en X , S está justificado en creer que x no es el/uno de los F .

En el caso de la paradoja del prefacio, éstas son:⁴³

1. El autor acepta que por lo menos una de las afirmaciones en su libro es falsa.
2. Para toda afirmación x en su libro, el autor acepta que esa afirmación x no es una de las las que son falsas.

Finalmente, en el caso de la lotería, éstas son:⁴⁴

1. Aceptamos que por lo menos uno de los boletos de la lotería ganará.
2. Para todo boleto x de la lotería, aceptamos que ese boleto x no es el que ganará.

Gracias a que tienen la misma estructura, las dos paradojas se resuelven de la misma manera, mostrando que de (1) y (2) no se sigue que el agente está siendo irracional o que está comprometido a aceptar una contradicción. Para ello, basta darse cuenta de que de lo que trata lo que el agente cree en (1) no es inconsistente con de lo que trata lo que cree en (2). Sin embargo, en este caso no basta hacer una distinción entre justificaciones generales y específicas ya que toda la paradoja opera solamente con información general, es decir, no apela a ninguna información sobre ningún boleto en específico. En otras palabras, en este caso no es posible decir, como en el caso de la paradoja del prefacio, que las premisas singulares no tratan sobre la lotería en general, sino sobre el boleto en particular al que refieren. Es necesario, pues, hacer distinciones más finas. Esta vez, es necesario

⁴³. Las X son las afirmaciones en el libro, y F es la propiedad de ser falsa.

⁴⁴. Los X son los boletos de lotería, y F es la propiedad de ganar la lotería.

considerar tres maneras distintas en las que alguien podría estar justificado en creer que su boleto de lotería no ganará: (a) porque está justificado en creer o sabe cuál es el boleto que ganará y no es el suyo, (b) porque está justificado en creer o sabe que otro boleto ganará, aunque no sabe cual, o (c) porque sabe o está justificado en creer que la lotería está amañada y no habrá ningún boleto ganador.⁴⁵ Ahora bien, lo interesante de estos tres tipos de justificación es que son incompatibles entre sí, es decir, si la justificación que tenemos para creer que un boleto no ganará es de un tipo, no es posible creer que otro boleto ganará con justificación de otro tipo. Para ser mas precisos, esto se debe a que (a) derrota las justificaciones (b) y (c), mientras que (c) derrota las justificaciones (a) y (b). Esto es así porque (a) y (c) son lógicamente inconsistentes entre sí (para que el boleto ganador sea distinto al nuestro es necesario que haya un boleto ganador), mientras que la justificación de tipo (b) es más débil que la de tipo (a) y (c): la justificación que nos dan (a) y (c) es absoluta, es decir, si sabemos cual es el boleto ganador y es distinto al nuestro o si sabemos que ningún boleto va a ganar, es posible afirmar con certeza que nuestro boleto no ganará; pero si sólo basamos nuestra creencia de que nuestro boleto no ganará en que el número de boletos es muy, muy grande, no es posible afirmarlo con la misma certeza. Esto significa que si tenemos la información necesaria para una justificación de tipo (a) o (c), esta información derrota nuestra posible justificación de tipo (b). Por ejemplo, si basé mi creencia de que mi boleto no ganará en mi creencia justificada de que la lotería que estoy jugando tiene 10,000,000 números, y luego me entero del número ganador y no es el mío, entonces mi primera justificación queda derrotada. Sigo creyendo que mi número no ganará, pero mi justificación es ahora otra, basada en conocimiento mas específico sobre el número ganador. Igualmente, si lo que me entero es mas bien que la lotería es una fraude y no habrá número ganador, entonces mi justificación original también queda derrotada, aunque mi creencia siga estando justificada. Lo que cambia es la fuente de su justificación.

Una vez que notamos esto, la solución a la paradoja aparece casi inmediata. El primer paso es preguntarse qué tipo de justificación tenemos para nuestra creencia en las premisas de la paradoja. No puede ser de tipo (a) porque no se aplica al número de lotería del cual tenemos justificación para creer que ganará.

⁴⁵. Por supuesto, puedo haber otras maneras en las que llegue a tener una creencia justificada de este tipo, pero considerarlas no enriquece mucho el análisis pues son todos casos que pueden tratarse de manera análoga a alguna de estos tres.

Entonces, debe ser de tipo de (b) o (c). Pero si es de tipo (c), entonces entonces el agente no tiene justificación para creer $\mathcal{N}+2$ y la paradoja se disuelve. Por otro lado, si la justificación de las premisas es de tipo (b), aunque sí es posible justificar así todas las premisas, esta justificación no puede extenderse a la conclusión $\mathcal{N}+3$, lo que también disuelve la paradoja.

Si algo aprendimos de los casos de conocimiento fácil, es que la justificación no se transmite a la presencia o ausencia de derrotadores. En otras palabras, si la justificación de nuestra creencia depende de que no se de un derrotador, entonces no es legítimo usar esa creencia para obtener justificación de que dicho derrotador se da o no. Esto se debe a que, por definición, el que nuestra justificación sea derrotada o no es algo externo a dicha justificación. En este caso, si la justificación de nuestras premisas es de tipo (b), es decir, está basada en nuestro conocimiento del número de boletos en la lotería y la resultante baja probabilidad de que un boleto dado gané, entonces el que no habrá ganador sería un derrotador de dicha justificación. Por lo tanto, aun cuando esta conclusión $\mathcal{N}+3$ se siga lógicamente de las premisas, nuestra justificación no se transmite a ella a través de dicha relación de consecuencia. La paradoja, por lo tanto, no se da, ya que quien cree justificadamente las premisas de 1 a $\mathcal{N}+2$ no está comprometido racionalmente a aceptar también $\mathcal{N}+3$.

E. El Sorites como Falla de Transmisión

Me parece que apelar así a de qué trata una creencia y cuales son sus derrotadores tiene muchas ventajas, entre ellas la de poder dar un diagnóstico y solución similar para paradojas y acertijos filosóficos que de otra manera parecerían muy disímiles, desde la paradoja de Moore y las ya conocidas fallas de clausura epistémica, hasta las ya mencionadas paradojas del prefacio y la lotería, y no quiero dejar de pasar la oportunidad de mostrar como apelar a condiciones de normalidad podría usarse para resolver otra famosa paradoja, aparentemente muy distinta las antes presentadas: la paradoja de *sorites*.⁴⁶

⁴⁶. Mi tratamiento de la paradoja sorites está inspirado en el trabajo de Delia Graff Fara (2000) y al igual que en el caso de la paradoja del prefacio y de la lotería, una versión muy preliminar apareció en Barceló (en prensa).

En una de sus versiones más conocidas, se dice que la paradoja se produce porque mientras el sentido común sugiere que los montones de arena tienen las siguientes propiedades, éstas son inconsistentes entre sí:

P. Dos o tres granos de arena no son un montón.

Q. Un millón (o cualquier número suficientemente grande) de granos de arena juntos sí son un montón.

R. Si n granos de arena no forman un montón, tampoco lo harán $(n+1)$ granos.

En otras palabras, si bien el sentido común sugiere aceptar como verdaderas estas tres proposiciones, de su conjunción se sigue una contradicción, a saber, (S) que un millón de granos de arena juntos son y no son un montón. Esto es un problema porque, idealmente, quisiéramos una teoría de los montones (y términos similares) que respetara nuestras intuiciones de sentido común sin comprometernos a aceptar una contradicción.

La solución de la paradoja empieza cuando nos preguntamos exactamente de qué tratan las premisas de la paradoja, y para responder esta pregunta es necesario preguntarnos en qué nos basamos para aceptarlas. Usando la distinción entre justificaciones específicas y generales que introduje unas páginas atrás, es muy probable que nuestra justificación de las primeras dos premisas sea específica a los grupos específicos de los que tratan: el de dos granos, el de tres y el de un millón. Considerando esos grupos en particular, y aplicando nuestra competencia con el término “montón”, concluimos que los primeros dos grupos no caen en su extensión, pero el tercero sí. Esta es la parte sencilla de mi solución, lo interesante sucede cuando nos preguntamos qué nos justifica a aceptar la tercera premisa, (R), es decir, que si n granos de arena no forman un montón, tampoco lo harán $(n+1)$ granos. En este punto es fundamental rescatar lo que dijimos unas páginas atrás sobre el uso de términos vagos, a partir del trabajo de Mario Gómez Torrente. Ahí dijimos que para ciertos usos de términos vagos,⁴⁷ es necesario que el contexto de uso sea uno en el que no haya series soríticas, es decir, ciertos usos de los términos vagos son derrotados por la existencia de series soríticas en el contexto de uso. Esto se debe a que, como he

⁴⁷. Agradezco a Ricardo Mena el haber llamado mi atención al hecho de que es posible usar los términos vagos en situaciones soríticas, es decir, que hay usos genuinos de términos vagos que no quedan derrotados por la existencia de una serie sorítica en su contexto. En estos casos, aun cuando la pretensión de hacer referencia alguna propiedad y/o denotar una extensión determinada queda derrotada, estos términos pueden seguir contribuyendo contenido semántico a lo que se dice. Diré más sobre cómo esto es posible un poco más adelante.

señalado ya, el uso de este tipo de términos no es infalible, es decir, no determina una extensión y/o propiedad determinada en cada contexto posible de uso, sino sólomente en aquellos en que se dan ciertas condiciones, como la de no haber una serie sorítica.

Para precisar exactamente de qué tipo de usos hablo y porqué dichos usos específicamente son derrotados por la existencia de series soríticas, permítaseme desarrollar un ejemplo simple. Supongamos que me encuentro parado al lado de la mesa del seminario, de tal manera que todos mis alumnos están sentados en fila, frente a mí. Ángel, Berenice, Carmen y Daniel están sentados en las cuatro sillas contiguas a mí, luego hay dos sillas vacías y en las que siguen están sentados Enrique y Felicia, la silla que sigue de Felicia está vacía y luego está sentado Gerardo en la silla al otro extremo de la mesa (donde, asumimos, cada silla está aproximadamente equidistante de la siguiente). Si quisiera hablar de los alumnos que están sentado “de mi lado”, la extensión de mi expresión quedaría perfectamente definida: serían Ángel, Berenice, Carmen y Daniel; y esto es así porque las dos sillas vacías entre Daniel y Enrique sirven como línea divisoria entre *mi lado* y *el otro lado*. En este sentido, el contexto en que uso “de mi lado” contribuye a determinar su extensión. Si mis estudiantes se hubieran sentado de manera distinta, la extensión de “de mi lado” sería diferente. Si Berenice y Carmen se cambiarán a ocupar los dos asientos vacíos entre Daniel y Enrique, la nueva extensión de “de mi lado” contendría sólomente a uno de mi alumnos, Ángel. Y en esta nueva configuración, Carmen dejaría de estar de mi lado, ¡*sin haberse movido!* Esto se debe a que “mi lado” no refiere a una ubicación espacial, sino que es un término que me permite, en contextos como el recién descrito, distinguir objetos por su locación espacial (relativa al hablante).

Ahora bien, supongamos que no fueran Berenice y Carmen, sino Felicia y Gerardo las que se movieran a ocupar los dos asientos entre Enrique y Daniel. Se formaría lo que en la literatura se conoce como una “Serie sorítica” y desaparecería el espacio que definía el límite entre estar de mi lado y estar del otro lado, y con él queda derrotado el uso del término para referir a un grupo de alumnos en este contexto. Esto quiere decir que usar “de mi lado” deja de servir para distinguir a mis estudiantes según su posición espacial relativa a la mía. Puedo seguir diciendo de algunos de mis alumnos – definitivamente de Ángel, y Berenice, y tal vez de Carmen – que están “de mi lado”, pero ya no puedo usar este término para referirme a un grupo con límites determinados. Este tipo de uso queda derrotado para este término. Así, cada vez que usamos un término vago para agrupar

objetos es necesario que haya, en el contexto de uso, un límite que distinga lo que cae bajo su extensión de lo que no.

Ahora estamos ya en posición de explicar qué nos justifica a creer la premisa (R) de la paradoja, a saber, que si n granos de arena no forman un montón, tampoco lo harán $(n+1)$ granos. Por principio de cuenta, podemos excluir fácilmente que nuestra justificación sea lo que unas páginas atrás he llamado una justificación específica. No llegamos a aceptar la premisa (R) porque pasamos por cada uno de los números, de uno a un millón, e inspeccionando dichos grupos terminamos concluyendo que o bien ambos eran montones o no lo eran. Tampoco proviene de una inferencia vacua, por ejemplo, de que no hay montones de granos de arroz, o que por el contrario, cualquier grupo de granos de arroz es un montón (pues ambas posibilidades son inconsistentes con las premisas P y Q). Mas bien, nuestra justificación es general, pues proviene de nuestra competencia con el término vago “montón” (y con la operación de añadir un grano a un montón). Así pues, cuando usamos el término vago “montón”, por ejemplo, para distinguir entre diferentes cantidades de granos de arroz, asumimos que en el contexto hay un límite entre lo que es un montón y lo que no es. Así pues, en dicho contexto, si es correcto decir de un grupo de n granos de arroz que no es un montón, lo es en parte porque hay en el contexto (por lo menos) un número mayor de n que cae dentro de un espacio vacío de posibles grupos de granos de arroz lo suficientemente grande y sobresaliente como para servir de límite entre los montones y los no-montones. En este contexto, $n+1$ no puede ser dicho número, porque entonces la diferencia no sería lo suficientemente grande y sobresaliente como para jugar dicho papel. Inferimos, correctamente por lo tanto, que el límite está también más allá de $n+1$, es decir, concluimos correctamente que el grupo de $n+1$ granos de arroz tampoco es un montón.

Sabemos (probablemente de manera inconsciente) que para que hablemos de un grupo de n granos de arroz que no es un montón, en el contexto en el que usamos “montón”, no para referirnos sólo al grupo en cuestión, sino para referirnos a un grupo con extensión determinada, debe haber un límite que separe los montones de los no-montones. Necesitamos algo análogo a las dos sillas vacías entre Daniel y Eduardo, pero con grupos de grupos de arroz. Necesitamos que en el contexto no se consideren todos los montones posibles con más de n granos de arroz, sino que haya un espacio vacío que sirva de límite entre los montones y los no-montones.

En su ausencia, el uso de “montón” quedaría derrotado y con él, nuestra justificación para creer que que si n granos de arena no forman un montón, tampoco lo harán $(n+1)$ granos. Es por ello que nuestra justificación de la premisa (R) quedaría derrotada por la existencia, en el contexto, de una serie sorítica. Sin embargo, esto es lo que sucede precisamente en la paradoja, se nos presenta con una serie sorítica y se espera que nuestra justificación de las premisas se mantenga, lo cual es imposible.

En otras palabras, de manera análoga a lo que sucedía en las paradojas anteriores, aceptamos que un número de granos no corresponde al límite entre lo que es un montón y lo que no lo es, asumiendo que dicho límite existe, sólo que se encuentra en *otro* lado. Cuando aceptamos que el límite no es la diferencia entre n y $n+1$, lo hacemos bajo la asunción de que *hay* un límite mas adelante o mas atrás. El error detrás de la paradoja sorítica es pensar que quién acepta que hay una línea divisoria entre lo que cae dentro de la extensión de un término y lo que cae fuera de él debe aceptar también de algún lugar que ahí se encuentra dicha línea divisoria. Espero haberlos convencido de que no es así, y que por lo tanto, nuestros juicios de sentido común no nos llevan a ninguna contradicción. (P), (Q) y (R) efectivamente son lógicamente inconsistentes entre sí, pero no podemos combiarlos en una inferencia que nos lleve a una contradicción, pues cada uno de ellos trata de algo distinto. (P) y (Q) tienen una justificación específica sobre los grupos particulares de 0 y un millón de granos, mientras que (R) tiene una justificación general que se derrota en situaciones soríticas. Así pues, aunque (R) es justificable en situaciones soríticas (por ejemplo, porque se sigue de manera vacua de que no hay montones o, por el contrario, que todo grupo de granos de arroz es un montón), esta justificación es inconsistente con la aceptación de (P) y (Q); mientras que la justificación de (R) no vacua no está disponible en contextos soríticos. De (R) por lo tanto, no se sigue (no en el sentido lógico, sino en el sentido de que su justificación no se transmite a) que si 2 granos de arroz no son un montón, tampoco lo son un millón de granos. Sin esta consecuencia, no se obtiene tampoco la conclusión contradictoria (S) de que un millón de granos de arroz son y no son un montón de granos de arroz.

En un texto previo (Barceló *en prensa*) había afirmado que el tratamiento que le doy a la paradoja sorítica muestra que ésta tiene poco que ver con el fenómeno de la vaguedad y mas bien tiene que ver con el fenómeno de la inferencia y, en particular, con la pregunta de cuando una inferencia lógicamente válida transmite la justificación epistémica de las premisas a la conclusión. Aun cuando creo haber mostrado que, efectivamente, la

paradoja sorítica, al igual que la del prefacio y la de la lotería, se disuelven una vez que nos damos cuenta que, dada la falibilidad de la justificación, ésta nunca se transmite a conclusiones que establecen la presencia o ausencia de derrotadores; sin embargo, ahora me parece un error decir que esto muestra que el sorites tiene poco que ver con la vaguedad. Ahora, preferiría adoptar una interpretación del sorites como la que ha defendido Miguel Ángel Fernández (2017) sobre la paradoja del escepticismo Cartesiano. Ahora prefiero decir que el sorites no plantea un desafío a nuestra comunicación lingüística ordinaria, de la misma manera “que otras paradojas filosóficas no constituyen desafíos para ninguna práctica o convicción ordinaria..., sino para la ambición filosófica que busca ofrecer un cierto tipo de teoría sistemática acerca de esas prácticas o convicciones.” (Fernández 2017, 73) En particular, constituye un desafío para las teorías de la comunicación lingüística que requieren que cada uso comunicativo exitoso de un predicado denote o fije una extensión o propiedad. Según lo que he sostenido en capítulos anteriores, este tipo de teorías ignoran o tergiversan el carácter falible de la comunicación lingüística y por eso no pueden dar cuenta del fenómeno de la vaguedad, como lo muestra la paradoja sorítica.

Resumiendo, si concebimos al lenguaje como una tecnología y a los términos lingüísticos como herramientas, como lo he propuesto en este libro, debemos asumir su falibilidad y sus consecuencias. Entre ellas, debemos aceptar la consecuencia de que nuestro uso de cualquier (o casi cualquier) término es derrotable. En particular, usar un término vago para referirse a un grupo determinado de objetos en un contexto es derrotable por la existencia de una serie sorítica en dicho contexto. Es por ello que, en la paradoja sorítica, aunque la conclusión se sigue de manera lógica de las premisas, no hay ningún contexto en el cual un agente puede estar justificado sobre las premisas y realizar la inferencia que transmita dicha justificación a la conclusión contradictoria. De esta manera, se disuelve la paradoja del sorites. En general, espero haber mostrado que introducir la noción de derrotabilidad a nuestra teoría de la justificación epistémica y llamar la atención a la cuestión de que ésta no se transmite a través de la relación de consecuencia lógica sino bajo ciertas condiciones nos ha servido para resolver tres conocidas paradojas filosóficas.

VII. Conclusiones

El tema central de este curso ha sido la falibilidad. Tal y cómo lo he caracterizado aquí, el fenómeno de la falibilidad surge de la relación entre lo determinado y lo indeterminado. El espacio en que esta relación cobra mayor importancia y, por lo tanto, donde más claramente se manifiesta la falibilidad es en nuestras prácticas normativas humanas. En el ámbito de lo normativo, la falibilidad surge del hecho de que comúnmente nos interesan cosas que escapan nuestro alcance y, sin embargo, no por eso dejamos de hacer un esfuerzo para conseguirlas. Cuando esto sucede, el esfuerzo que hacemos, por más cuidadoso, responsable, etc. que sea, muchas veces (tal vez todas) no puede sino ser falible, es decir, no llega a garantizar el éxito de nuestra empresa. En otras palabras, muchas veces, nos planteamos lo que aquí he llamado valores “trascendentes” – valores que escapan nuestro alcance, como el evitar dañar a los otros, decir cosas verdaderas, referirnos a cosas que de hecho existen, hacer reír a la gente, etc. – alrededor de los cuales desarrollamos prácticas normativas que plantean otro tipo de valores. Valores de este segundo tipo – que he llamado “instrumentales” – regulan nuestras decisiones de cómo actuar, cómo evaluar nuestros actos y los de otros, cómo reaccionar ante las consecuencias de nuestros actos, etc. de una manera que está relacionada con nuestros fines trascendentes, pero también los trasciende, vélgase la redundancia.

La falibilidad es el hecho de que actuar siguiendo nuestras reglas y valores instrumentales no garantiza que alcancé nuestros fines trascendentes; y vice versa, el que alcancé nuestros fines trascendentes o no, no determina por completo si actuamos con el suficiente cuidado y responsabilidad, es decir, cumpliendo nuestras obligaciones, reglas y valores instrumentales. En otras palabras, es posible hacer todo lo que está en nuestras manos para actuar de manera cuidadosa y responsable y aun así es posible que nuestros actos tengan consecuencias negativas; igualmente es posible que nuestros actos tengan consecuencias positivas sin que esto signifique que he sido cuidadoso o negligente. Cuando estos dos tipos de valores divergen, decimos que tuvimos buena o mala suerte. El que actúa con cuidado y aun así no logra consecuencias positivas tuvo mala suerte. El que actúa de manera descuidada y negligente, pero sus actos no tienen consecuencias negativas, en contraste, tiene buena suerte. Finalmente, hay también casos más raros, conocidos como “casos de Gettier”

donde los medios son los adecuados, se alcanzan los fines y, sin embargo, los fines se alcanzan por (buena) suerte y no debido a que los medios fueron los adecuados. y vice versa, hay casos donde los medios no son los adecuados, se falla en alcanzar los los fines y, sin embargo, los fines no se alcanzan por (mala) suerte y no debido a que los medios fueron inadecuados.

En general, he ofrecido respuesta a dos preguntas esenciales alrededor de esta distinción entre hacer el esfuerzo correcto para alcanzar una meta, alcanzar la meta y lograr alcanzar la meta (es decir, alcanzarla precisamente porque se hizo el esfuerzo correcto): la cuestión *descriptiva* de explicar en qué consiste exactamente el que lograr un objetivo se deba al ejercicio de alguna capacidad o sea el resultado directo de hacer un esfuerzo para alcanzarlo, es decir, explicar exactamente qué tipo de relación debe haber entre el esfuerzo o ejercicio de una capacidad y la obtención del objetivo para decir que alcanzar dicho objetivo fue un logro y no un afortunado accidente, y la cuestión *fundacional* de porqué valoramos más un logro (debido al esfuerzo propio o al ejercicio de alguna capacidad) que el obtener lo que deseamos de manera accidental o por buena suerte, mientras que castigamos igual los casos en los que el mal es resultado de mala suerte que cuando no lo es.

Para dar respuesta a la pregunta descriptiva he desarrollado una noción técnica de previsibilidad, donde el que una posibilidad abierta sea ‘lo suficientemente previsible’ significa que sería irresponsable no prepararse adecuadamente para garantizar el éxito de la empresa en caso de que dicha posibilidad se actualizara. La idea básica es muy simple: cuando tenemos que actuar en situaciones de recursos limitados en pos de un fin que nos trasciende en el momento, el problema es que aun quedan abiertas muchas posibilidades que pueden afectar el que alcancé nuestro fin o no y qué debé hacer para lograrlo. Es exagerado esperar estar preparados para cualquier posibilidad abierta, y basta estar preparados solamente para algunas: aquellas que son lo que he llamado “suficientemente previsible.” Según mi teoría, lo que hace que una situación abierta (a la hora de actuar o evaluar) sea lo *suficientemente previsible* en este sentido no es solamente qué tan fácil sería anticipar que podría suceder sino que es una combinación de varios factores como :

1. Qué tan probable es la situación, dado lo que el agente sabe
2. Qué tan costoso es prepararse adecuadamente para garantizar el éxito en el caso de que efectivamente se de dicha posibilidad

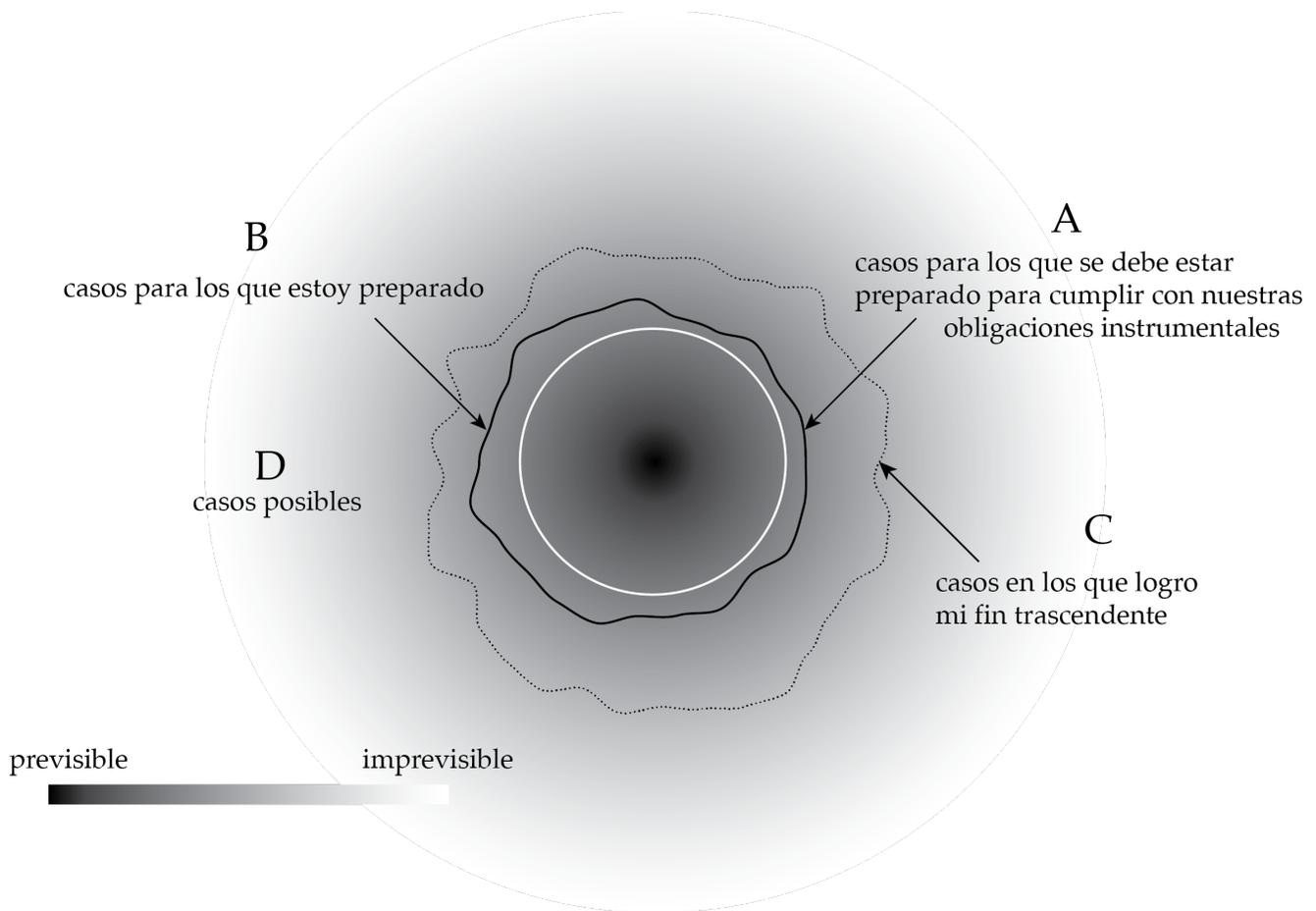
3. Qué diferencia haría el que se tomará en cuenta dicha posibilidad o no

4. Qué tan claro sería para los otros que efectivamente uno se preparo adecuadamente para dicha situación

5. Convenciones operables en el contexto sobre qué posibilidades deben tomarse en cuenta y cuales no

6. Otras contingencias históricas.

Ahora bien, cuando un agente⁴⁸ esta justificado, debé distinguir los siguientes cuatro diferentes tipos de situaciones posibles. Dado que la propiedad de ‘ser previsible’ es gradual, cada uno de estos tipos es mas previsible que los siguientes:



A. Posibilidades para las que necesitamos garantizar el éxito para estar justificados

B. Posibilidades para las que de hecho el agente ha garantizado el éxito

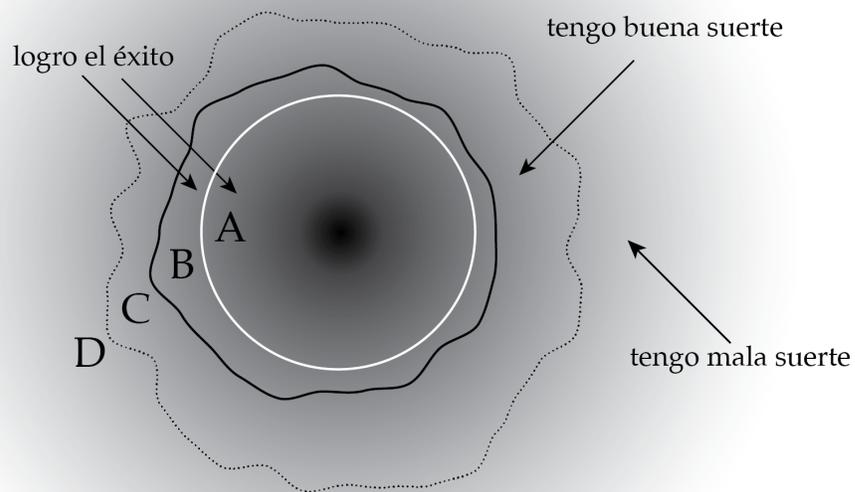
C. Posibilidades en las que habríamos alcanzado nuestro fin

⁴⁸. Como he señalado a lo largo del libro, en realidad no es necesario siquiera que tengamos aquí un agente, pues el fenómeno es mucho mas general.

D. Posibilidades abiertas

En el caso de la justificación epistémica, estos cuatro niveles corresponden a:

- A. Posibilidades para las que necesitamos garantizar la verdad para estar justificados
- B. Posibilidades para las que de hecho el agente ha garantizado la verdad de su creencia
- C. Posibilidades en las que la creencia es verdadera
- D. Posibilidades abiertas



Cuando el agente está justificado, cada uno de estos conjuntos de posibilidades está contenido en el que le sigue, es decir, *D* contiene a *C* que contiene a *B* que contiene a *A*. El que *A* sea distinto de *C* es lo que explica que la justificación, en general sea factible; el que *B* sea distinto de *C* explica porque esta justificación en particular de este agente es falible.

La pregunta por la suerte sólo surge respecto a cual posibilidad se actualiza:

Si la posibilidad que se actualiza está en

- B*, entonces el agente tuvo éxito
- C* (pero no en *B*), entonces el agente tuvo buena suerte
- D* (pero no en *C*), entonces el agente tuvo mala suerte

Con esta distinción un mano, es posible explicar la distinción entre un acierto afortunado y un verdadero logro de la siguiente manera. Un agente S logra su fin F a través de un medio M si:

(1) El agente alcanza F

(2) El medio M era adecuado para conseguir F dada la situación del agente al momento de actuar, es decir, si y sólo si al momento de actuar el agente S estaba adecuadamente preparado para garantizar el éxito para toda posibilidad abierta (en su contexto de acción) suficientemente previsible.

(4) y el agente S alcanzó su fin F porque el medio M que usó para alcanzarlo era adecuado para este fin, es decir, si de hecho se dio una de las situaciones para las que el agente estaba adecuadamente preparado (es decir, se actualizo una situación dentro de B).

En este curso he defendido un pluralismo de valores que he llamado “dualismo” porque fundamente distingue entre dos tipos de prácticas normativas: En primer lugar, tenemos la práctica de adoptar fines o valores trascendentes, es decir, que dependen de factores que no están bajo nuestra disposición y/o control a la hora de actuar. Estos valores establecen los fines de nuestro actuar y a partir de ellos es posible evaluar nuestros medios dependiendo de qué tanto nos acercan o no a nuestros fines. Para el consecuencialista, este tipo de normatividad es la normatividad fundamental (si se es un consecuencialista moderado) o el único tipo genuino de normatividad (si se es un consecuencialista radical). Este tipo de normatividad suele manifestarse ‘hacia adelante’, es decir, *ex post facto*: dado que comúnmente toma tiempo determinar si el acto alcanzó su objetivo o no (o si tuvo consecuencias positivas o negativas), antes de eso sólo es posible hacer predicciones mas o menos racionales de si se alcanzará el objetivo o no. El acto no es bueno o malo sino hasta que tiene las consecuencias relevantes. Es entonces que es posible actuar normativamente sobre el agente dependiendo de las consecuencias de sus actos. En este sentido, ser responsable tiene que ver, no sólo con lo que el agente hace al momento de actuar o antes, sino también con lo que hace después, a cómo reacciona ante las consecuencias de sus actos, pidiendo perdón, haciendo enmiendas de sus errores, etc. Desde el punto de vista legal, esto justifica prácticas asociadas al concepto de *tort* como son las demandas, indemnizaciones, compensaciones, etc.

En contraste y en segundo lugar, tenemos la práctica de tomar decisiones en situaciones de indeterminación y recursos limitados. En estos casos, necesitamos valores y reglas cuya aplicación dependa sólo de aquello que tenemos nuestra disposición y/o control. Sólo este tipo de valores y reglas pueden guiarnos en nuestra toma de decisiones. Para el deontologista, este tipo de normatividad es la normatividad fundamental (si se es un deontologista moderado) o el único tipo genuino de normatividad (si se es un deontologista radical) ya que es el único que depende del agente y, por lo tanto, el único en el que evaluamos el acto *qua acto*. Este tipo de normatividad suele manifestarse ‘hacia atrás’, es decir, *ex ante facto*, en tanto depende solamente de factores como las intenciones del acto, el cuidado que se tuvo al actuar, la preparación que se hizo, el comportamiento anterior en casos similares, etc., es decir, de factores que quedan determinados a la hora de actuar o antes. Por ello, es posible actuar normativamente sobre el agente inmediatamente después del acto (y hasta antes, por ejemplo, cuando se clausura un restaurante poco-higiénico aun antes de causar una infección o intoxicación en algún parroquiano) aun antes de determinar sus consecuencias, pero no después en base a factores que no estaban disponibles o bajo el control del agente a la hora de actuar. Es por ello que solé decir cosas como que “el hubiera no existe”. Desde el punto de vista legal, esto justifica prácticas preventivas como la emisión de licencias, inspecciones, reglamentos, etc.

En vez de adoptar un tipo de monismo normativo donde sólo uno de estos dos tipos tan distintos de practicas normativas es genuinamente normativo, o un dialeteísmo que simplemente acepta que a veces hay cosas que son al mismo tiempo y en el mismo sentido correctas e incorrectas, he ofrecido una alternativa pluralista en la cual ambas practicas son correctas en su propio ámbito de aplicación. Este dualismo es muy cercano al relativista propuesto por John MacFatlane y por eso he dedicado gran parte del curso a distinguir mi posición de la suya.

En general, siempre que tenemos razones encontradas respecto a si una entidad E tiene una propiedad P , es posible adoptar una de seis posiciones: **monismos** (1) que E efectivamente es P , y que nuestras razones a favor de que E no es P están equivocadas, (2) que en realidad E no es P , y que nuestras razones a favor de que E es P están equivocadas, **dialeteísmo** (3) que E es tanto P como no P y que nuestro miedo a las contradicciones está infundado, **gradualismo** (4) que P es una propiedad gradual y E es un caso intermedio, **relativismo** (5)

que hay un tercer elemento involucrado C tal que E es P de acuerdo a algunos C s y no P desde otros, **dualismo de entidades** (6) que donde pensábamos que existía sólo una entidad E , en realidad hay dos: E_1 y E_2 tales que E_1 es P y E_2 no, **dualismo de propiedades** (7) que donde pensábamos que existía sólo una propiedad P , en realidad hay dos: P_1 y P_2 tales que E es P_1 y no P_2 .

Cada una de estas posiciones tiene sus retos y virtudes. El primer tipo de monismo tiene que mostrar porqué tenemos razones a favor de que E no es P pese a que es falso, mientras que nuestras razones a favor de que E es P sí son confiables; el monismo contrario tiene que mostrar lo contrario. El dialeísmo tiene que mostrar cómo es posible que haya contradicciones genuinas en el mundo. El gradualismo tiene que mostrar cómo es posible que algo no sea ni P ni no- P . Contra el monismo, el relativismo debe mostrar cómo, el que entidades como E tengan propiedades como P efectivamente depende del factor C ; contra el dualismo y el gradualismo, el relativismo debe mostrar que C es un aspecto genuino de la realidad y no algo que se ha postulado ad-hoc para conciliar nuestras razones en conflicto, es decir, debe mostrar que la existencia de C se manifiesta de otras maneras. Los dualismos a su vez deben mostrar que las distinciones que postulan no son ad-hoc sino que su postulación sirve para dar cuenta de otros fenómenos independientes. En contra del relativismo, deben argumentar que ese tercer elemento que los relativistas postulan no es autónomo sino que sólo es una manera de determinar los verdaderos objetos de los que se predica el predicado o las verdaderas propiedades que se le predicán al objeto, y que no tiene otra función más que ésa.

En mi caso particular, he tratado de mostrar que la postulación de un tercer elemento – lo que MacFarlane llama la perspectiva de evaluación – que hace el relativista, aunque superior a los monismos deontologista y consecuencialista, no logra dar cuenta cabalmente de la complejidad de nuestras prácticas normativas. En particular, no logra dar cuenta de cómo los tres tipos de práctica normativa que he identificado – que generalizan la distinción que se hace en epistemología entre verdad, justificación y conocimiento – son tan sustancialmente distintos y no son, como quisiera el relativista, meras manifestaciones distintas de una misma práctica normativa desde diferentes perspectivas.

REFERENCIAS

- Allen, Colin (2010) “Mirror, Mirror in the Brain, What’s the Monkey Stand to Gain?”, *Nous*, 44:2, 372–391
- Alston, J. (2000), *Illocutionary Acts and Sentence Meaning*. Ithaca: Cornell University Press.
- Asse, Jacobo (2015) *Una teoría artefactual de los objetos matemáticos*, tesis de doctorado en filosofía de la ciencia, UNAM.
- Austin, J. L. (1956, 1957) “Un alegato en pro de las excusas” en *Ensayos filosóficos*, compilados por J. O. Urmson y G. J. Warnock; traducción y presentación de Alfonso García Suárez, Madrid: Revista de Occidente, 1975
- Barceló, Axel (manuscrito) “Extensión y Blanco de los Epítetos”
- (2017) “Inferencia y Presuposición”, manuscrito bajo dictamen.
- (2015) “Making Quantified Truths True”, en Torza, Alessandro (Ed.) *Quantifiers, Quantifiers, and Quantifiers: Themes in Logic, Metaphysics, and Language*, Synthese Library, 323-338
- (2012) “Semantic and Moral Luck”, *Metaphilosophy*. Volumen 43, Número 3, Abril 2012, pp. 204-220
- (2002) “Universabilidad y Aplicabilidad de las Matemáticas en Wittgenstein y el Empirismo Logicista” *Theoría, Revista del Colegio de Filosofía*. México: Facultad de Filosofía y Letras, UNAM. Núm. 13, Junio 2002. Pp.119-136.
- Barceló, Axel y Robert Stainton (2018) “Quasi-Factives and Cognitive Efficiency”, en *Relevance: Pragmatics and Interpretation*, editado por Kate Scott, Billy Clark y Robyn Carston, Cambridge University Press.
- Bar-Eli, Michael; Ofer H. Azar1 y Yotam Lurie (2009) “(Ir)rationality in action: do soccer players and goalkeepers fail to learn how to best perform during a penalty kick?”, en M. Raab et al. (eds.) *Progress in Brain Research*, vol. 174, 97-108.
- Barwise, Jon & John Perry, (1983) *Situations and Attitudes*, MIT Press.
- Berker, Selim. (2013) “The Rejection of Epistemic Consequentialism.” *Philosophical Issues* 23 (1) (October): 363–387
- Brandom, Robert (1994), *Making it Explicit*. Cambridge: Harvard University Press.

- Bortolotti, Lisa (2016) “The Ethics of Delusional Belief” *Erkenntnis*, April 2016, Volume 81, Issue 2, pp 275–296
- Come S. (2002), “Basic Knowledge and the Problem of Easy Knowledge,” *Philosophy and Phenomenological Research*, 65.2: 309–329.
- Deleuze, Gilles (2001) “Presentación de Sacher-Masoch.” *Lo frío y lo cruel*, Harvard: 84-93.
- Dennet, Daniel (1990) “The interpretation of texts, people and other artifacts”, *Philosophy and Phenomenological Research*, 177-194.
- Descartes, René (1641) *Meditaciones metafísicas con objeciones y respuestas*, trad. de Peña García, Vidal. Alfaguara, Madrid, 1993
- (1637) *Discurso del método para bien dirigir la razón y buscar la verdad en las ciencias*, introducción de Miguel Angel Granada. Traducción y notas de Eugenio Frutos. Ed. Planeta, Barcelona, 1984, 205 págs.
- Doris, J. M. (2002) *Lack of character: Personality and moral behavior*, Cambridge University Press.
- Douven, I. (2006), “Assertion, knowledge, and rational credibility”, *Philosophical Review* 115: 449–485.
- Dowell, Janice (manuscrito) “When are Truth-assessments Data for Semantic Theorizing about Epistemic Modals?”
- Dutant, Julien (2015), “The legend of the justified true belief analysis.” *Philosophical Perspectives*, 29: 95–145. doi: 10.1111/phpe.12061
- Earl, Dennis (2017) “Concepts” en *The Internet Encyclopedia of Philosophy*, ISSN 2161-0002, <http://www.iep.utm.edu/>, retrieved on January 8th, 2017.
- Elder, C. (2007) “On the place of Artifacts in Ontology” en Eric Margolis y Stephen Laurence (eds.): *Creations of the Mind: Theories of Artifacts and Their Representation*, Oxford University Press, 33-51.
- Edmonds, M. Leslie (1987) “Superstition to Science: The Role of Research in Strengthening Public Library Service to Children”, *Library Trends*, 35 (3): 509-520.
- Eraña, Ángeles (2003) “Normatividad epistémica y estructura heurística del razonamiento”, *Crítica, Revista Hispanoamericana de Filosofía*. Vol. 35, No. 104 (agosto 2003): 69–108
- Eraña, Ángeles y Axel Barceló (2016) “El conocimiento como una actividad colectiva”, *Tópicos, Revista de Filosofía* 51: 9-35.

- Fara, Delia Graff (2000) “Shifting Sands: An Interest-Relative Theory of Vagueness”, *Philosophical Topics* 28(1): 45–81.
- Fernández, Miguel Ángel (2017) “El desafío del escepticismo cartesiano”, *teorema* 36(2): 69-104.
- (2013) “The Possibility of Epistemic Responsibility”, *Philosophical Issues*, 23, *Epistemic Agency*, pp. 110-131.
- Franssen, Maarten (2006) “The Normativity of Artefacts”, *Studies in History and Philosophy of Science*, 37: 42-57.
- García-Aguilar, Claudia Lorena (2004) “Especificidad de dominio y sesgo en la psicología del razonamiento humano”, *Signos Filosóficos*, vol. VI, núm. 12, julio-diciembre, 2004, pp. 63-91
- Gardner, John (2007) *Offences and Defences*, Oxford University Press.
- Gómez-Torrente, Mario, (2010), “The Sorites, Linguistic Preconceptions, and the Dual Picture of Vagueness”, en R. Dietz y S. Moruzzi (eds.), *Cuts and Clouds. Essays in the Nature and Logic of Vagueness*, Oxford University Press, Oxford, 228-253.
- Grandy, R. (2007) “Artifacts: parts and principles” en Eric Margolis y Stephen Laurence (eds.): *Creations of the Mind: Theories of Artifacts and Their Representation*, Oxford University Press, 18-32.
- Greco, John (2010) *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*, Cambridge University Press.
- Green, Mitchell S. (1999) “Attitude Ascription’s Affinity to Measurement”, *International Journal of Philosophical Studies* 7(3): 323-348.
- Hansson, Ove; Karin Edvarsson Björnberg y John Cantwell (2016) “Self-Defeating Goals”, *Dialéctica*, Volume 70, Issue 4, December 2016, Pages 491–512
- Harman, Gilbert (1999) “Moral Philosophy Meets Social Psychology”, *Proceedings of the Aristotelian Society*, New Series, Vol. 99 (1999), pp. 315-331
- Haslanger, Sally (2000) “Gender and Race: (What) Are They? (What) Do We Want Them To Be?”, *Nous* 34(1): 31-55.
- Hernández, Héctor (2012) *Las Fascinantes Paradojas de la Probabilidad: Concepto, resolución e importancia en las ciencias y en la filosofía*, Editorial Académica Española.

- Hetherington, Stephen (2017) “Fallibilism”, *The Internet Encyclopedia of Philosophy*, ISSN 2161-0002, <http://www.iep.utm.edu/>, consultada el 7 de enero.
- Kvanvig, J.(2005), ‘Truth is not the Primary Epistemic Goal’, in *Contemporary Debates in Epistemology*, eds. E. Sosa and M. Steup, Oxford: Blackwell, pp. 285–96.
- Korsgaard, Christine M. et.al. (2000) *Las fuentes de la normatividad*, traducida por Laura Lecuona y E. Manríquez, traducción revisada por Faviola Rivera, Instituto de Investigaciones Filosóficas, Colección:Filosofía Contemporánea, Ciudad de México.
- Kroedel, Tomas (2013) “The Lottery Paradox, Epistemic Justification, and Permissibility”, *Logos & Episteme* 4(1): 103-111.
- Kroes, P. (2012) *Technical Artifacts: Creations of Mind and Matter*, Springer: Holanda.
- Langton, Rae & Caroline West (1999) “Scorekeeping in a pornographic language game”, *Australasian Journal of Philosophy*, 77:3, 303-319
- Littlejohn, Clayton (*en prensa*) “The Unity of Reason”, en John Turri (ed.), *Epistemic Norms: New Essays on Action, Belief, and Assertion*.
- (2013) “Don’t know, don’t believe: reply to Kroedel”, *Logos & Episteme* 4(2): 231-238.
- MacFarlane, John (2014). *Assessment Sensitivity: Relative Truth and its Applications*. Oxford University Press.
- Mackenzie, Donald (2001) *Mechanizing Proof: Computing, Risk, and Trust*, The MIT Press.
- Magallanes, Alejandro (2011) *¿Con qué rima tima?*, Almadía.
- Martin, Justin W. & Fiery Cushman (2015) “The Adaptive Logic of Moral Luck”, en Justin Sytsma y Joshua Knobe (eds.) *Blackwell Companion to Experimental Philosophy*, Blackwell.
- Moi, Toril, (1999) *What is a Woman? and Other Essays*, Oxford: Oxford University Press.
- Montemayor, Carlos (2014) “Success, Minimal Agency and Epistemic Virtue”, en Abrol Fairweather (ed.): *Virtue Epistemology Naturalized: Bridges Between Virtue Epistemology and Philosophy of Science*, Synthese Library 366, Springer, 67-81
- Morado, Raymundo (2000) “Construcción, Reconstrucción y Evaluación de Argumentos”, apuntes del 26 de mayo del 2000 para el curso *Filosofía del Lenguaje (Análisis Lógico de Argumentos)*, Colegio de Filosofía,

Facultad de Filosofía y Letras. Julio a Noviembre del 2000, Grupo 0021, clave 0841 (3416 para estudiantes del plan 1999). Consultado en línea el 13 de febrero de 2017 en <http://www.filosoficas.unam.mx/~morado/Cursos/0Argumentos/cons.htm>

Morton, Adam (2006) Review of Alan Miller's *Understanding People, Mind* 115 (459) 780

Nagel, Thomas, (1979), "Moral Luck", en *Mortal Questions*, Nueva York, Cambridge University Press. Traducción al castellano: *La muerte en cuestión*, México, FCE, 1981. Se ha reeditado con el título *Ensayos sobre la vida humana*, 2000.

Nozick, R. (1993), *The Nature of Rationality*, Princeton, NJ: Princeton University Press.

Padilla, Hugo (1971) "Condiciones para saber. (Sobre un artículo de Edmund L. Gettier)", *Crítica, Revista Hispanoamericana de Filosofía*, 5(13): 129-134.

Peacocke, C. (2005). Rationale and Maxims in the Study of Concepts, *Noûs*, 39 (1): 167-78.

Peirce, Charles Saunders (1868) "Some Consequences of Four Incapacities." *The Journal of Speculative Philosophy* 2(3):140-157.

Posner, Richard A. (1972) "A Theory of Negligence", *The Journal of Legal Studies*, 1(§1): 29-96.

Priest, Graham (1985) "Inconsistencies in motion", *American Philosophical Quarterly* 22, no. 4: 339-346.

Rawls, John (1955) "Two Concepts of Rules", *The Philosophical Reviews*, 64: 3-32.

Rescorla, Michael (2009), "Assertion and its Constitutive Norms", *Philosophy and Phenomenological Research*, Vol. LXXIX No. 1, July 2009, pp. 98 - 130.

Rosell, Sergi (2006) "Nagel y Williams acerca de la suerte moral", *Revista de Filosofía* 1(1): 143-165

Rudy, Fernando (manuscrito)

Ryle, Gilbert (1953) "Ordinary Language", *The Philosophical Review*, Vol. 62, No. 2 (Apr., 1953), pp. 167-186

Salmon, Nathan (1986) *Frege's Puzzle*, Cambridge, Mass. The MIT Press.

Sartorio, Carolina, (2015) "Resultant Luck and the Thirsty Traveler", *Methodos* 6, 153-171.

Searle, John (1969), *Speech Acts*. Cambridge: Cambridge University Press.

(2001), *Rationality in Action*. Cambridge: MIT Press.

- Sellars, Wilfred (1963), "Some reflections on language games", reimpresso en *Science, Perception, and Reality*. New York: Routledge and Kegan.
- Shiota, Michelle N. (manuscrito) "'Fame' is the Problem: Conflation of Visibility With Potential for Long-Term Impact in Psychological Science."
- Singal, Jesse (2017) "It Is Really Important to Humanize Evil", *New York Magazine / Science of US*, publicado y leído el 19 de Mayo, 9:48 am. URL = <http://nymag.com/scienceofus/2017/05/it-is-really-important-to-humanize-evil.html?mid=fb-share-scienceofus>
- Sperber, Dan (2007) "Seedless Grapes: Nature and Culture" en Eric Margolis y Stephen Laurence (eds.): *Creations of the Mind: Theories of Artifacts and Their Representation*, Oxford University Press, 124-137.
- Steglich-Petersen, Asbjørn (2006) "No Norm Needed: on the Aim of Belief", *The Philosophical Quarterly* 56 (225), 499-516.
- Strawson, Peter Frederick (1959) *Individuals: An Essay in Descriptive Metaphysics*. Londres, Methuen.
- Travis, Charles (1985) "On What Is Strictly Speaking True", *Canadian Journal of Philosophy*, 15 (2):187 - 229.
- van Eemeren, Frans H. and Houtlosser, Peter, (2003) "More about Fallacies as Derailments of Strategic Maneuvering: The Case of Tu Quoque" Ossa Conference Archive. 93.
- Velleman, D. (2000) *The Possibility of Practical Reason*, Oxford University Press.
- Williams, Bernard (1981): "Moral Luck", en *Moral Luck*, Cambridge, Cambridge University Press. Traducción al castellano: La fortuna moral, México, UNAM, 1993.
- Williamson, Timothy (2000) *Knowledge and its Limits*. Oxford: Oxford University Press.
- Yablo, Stephen (2014) *Aboutness*, Princeton University Press.
- Yoder, Robert M. (1951) "Someday They'll Get Slick Willie Sutton", *The Saturday Evening Post*, 223(30), 17, January 20, Saturday Evening Post Society, Indianapolis, Indiana.
- Zheng, Robin y Nils-Hennes Stear (manuscrito) "Imagining in oppressive contexts, or, what's (intrinsically) wrong with 'Blacking Up'?"
- Zuluaga, Mauricio (2005) "El Problema de Agripa", *Ideas y Valores* 54 (28): 61-88.

