

# Fallar es humano

## un análisis filosófico de la falibilidad

Axel Arturo Barceló Aspeitia

[abarcelo@filosoficas.unam.mx](mailto: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) Traducción mía.<sup>1</sup>

El objetivo de este curso es analizar la estructura de la falibilidad a través de diferentes campos normativos de interés filosófico y en especial, la epistemología, la ética y el lenguaje. Al terminar el curso, habremos construido lo que podríamos llamar una teoría general de la normatividad dibujando las relaciones entre una serie de conceptos normativos fundamentales como los de falibilidad, suerte, virtud, responsabilidad, preparación, negligencia, juicio, y similares. Trabajaremos bajo la hipótesis de que esta estructura normativa es la misma en todos los campos normativos y lo ilustraremos con ejemplos provenientes de los tres campos ya mencionados y otros más. Si tenemos éxito, podremos dar cuenta de fenómenos tan aparentemente disímiles como como prepararse de manera responsable para un desastre natural hasta cómo comunicarse con lenguaje vago. Lo que no haremos será ofrecer una explicación de la normatividad en términos no

---

<sup>1</sup>. “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.”

normativos. Dudo que la normatividad no sea irreducible, pero esa no será una de las cuestiones que abordaremos. Otro monstruo que dejaremos dormir a nuestro lado será el del escepticismo. En otras palabras, no diremos nada sobre cómo “responder al reto escéptico”, lo que sea que eso signifique. Mas bien, trabajaremos bajo el supuesto de que el discurso normativo tiene sentido, es decir, que podemos actuar de manera responsable o irresponsable, que el éxito y el fracaso son ambos posibles y que no es lo mismo hacer las cosas bien que mal. Muchos filósofos han presentado argumentos muy brillantes contra de estos supuestos, pero no los abordaremos aquí.

El curso estará estructurado alrededor de seis conceptos centrales del fenómeno normativo:

## **I. Falibilidad**

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, podemos definir un proceso como falible, en general, si su éxito no está garantizado, es decir, si efectuar el proceso de manera responsable e irreprochable no garantiza que se alcance el objetivo. El objetivo de este curso 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 toda creencia humana está justificada de manera falible. Sin embargo, me parece claro que la falibilidad epistémica es sólo un caso de un fenómeno mas general que puede darse en otros ámbitos. 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.

### **A. Suerte**

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 irreprochable, sino por suerte. En esta segunda sección estudiaremos 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.

## **B. Virtudes**

¿Qué distingue un caso de éxito genuino de uno que se da solo por suerte? En esta tercera sección veremos una propuesta reciente de respuesta según la cual para que alcanzar un objetivo cuente como un verdadero logro es necesario que manifieste cierta competencia para alcanzar dichos logros.

## **C. Relativismo**

En esta cuarta sección abordamos la cuestión de ¿en qué sentido puede ser efectivamente irreprochable un proceso falible que no resulta en aquello que persigue? Esto significa, en el caso epistémico, preguntarse si realmente hay algo que podamos evaluar positivamente en un caso de creencia falsa justificada o, en ética, si podemos evaluar como moralmente correcto un acto que se realice de manera responsable pero tuvo consecuencias excesivamente dañinas para una víctima inocente. En particular, abordaremos la propuesta relativista según la cual nuestra evaluación de un proceso falible varía conforme vamos teniendo mayor o menor certeza de que alcanzara su éxito o no. En este sentido, lo correcto o incorrecto de un acto puede cambiar conforme va produciendo consecuencias positivas o negativas.

## **D. Dualismo**

En la quinta sección propondré otra respuesta a esta pregunta, según la cual es necesario distinguir dos diferentes tipos de normatividad – una dirigida hacia los fines y otra a los medios. Para ello, revisaremos evidencia empírica de que, efectivamente, contamos con dos sistemas cognitivos de evaluación relativamente independientes los cuales, en casos normales, no generan resultados divergentes y que, por eso, los concebimos como operando sobre un solo dominio normativo.

## **E. Presuposición y derrotabilidad**

En la última sección estudiaremos 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*. Ilustraré la

importancia de estos presupuestos mostrando como dan pie a algunas de nuestras paradojas más conocidas como la del prólogo o la de sorites.

## II. Falibilidad

### 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 Durant (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, pensemos 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 bien la razón no es más que buscar la verdad. En otras palabras, el único objetivo es la verdad y lo que debemos 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 más sobre esto más adelante. Entonces pensemos 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 más 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 más 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 por qué a veces cometemos 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.

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 esta 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 sabemos que no basta señalar al mal y decir que es malo. Hay que hilar más 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 falibilista es difícil hacer sentido de nuestra práctica ordinaria 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 más 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 conclusión 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, hacemos 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 podemos 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 hacemos, 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 podemos dejar a nuestras prácticas evaluativas ordinarias guiarnos hacia un mejor entendimiento de la normatividad.

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 quienes 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 concept 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 mas 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 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 podemos 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 consideremos ejemplos en otros campos normativos: el contraste entre un método (herramienta, procedimiento, mecanismo, acto, 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 presentaremos 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 proponemos algo, no hacemos todo lo que podemos 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 detendremos 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 pensemos 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 creemos. 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 valoremos 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

Hace unos meses, tuve la oportunidad de escuchar a Miriam Schoenfield decir algo muy obvio, pero no por ello poco profundo, sobre la epistemología. Decía que, si uno valora la verdad, para la pregunta ¿qué debemos creer? había una respuesta simple y obvia: debemos creer lo que es verdad. Sin embargo, si la pregunta de ¿qué debemos creer? se respondiera tan fácilmente, los libros de epistemología serían mucho, mucho mas cortos. Para entender porqué, permítaseme poner otro ejemplo completamente análogo pero mas obvio y hasta gracioso. Supongamos que alguien presumiera saber la manera de hacerse rico en la lotería, es decir, qué debemos hacer para hacernos ricos jugando la lotería y le preguntaremos cual 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, haremos muchos 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, mas bien, una definición de *lo que es* ganar la lotería. En otras palabras, lo que pedimos fue un medio, y lo que se nos dió fue (una re-descripción de) la meta.

Lo mismo sucede con la pregunta ¿qué debemos creer? Responder que debemos 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 debemos creer de lo que no. Al igual que en el caso de la lo 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.

Comparemos dos posibles reglas epistémicas: “debemos creer cosas verdaderas”, “debemos creer sólo aquello de lo cual no podemos 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, “debemos creer sólo aquello de lo cual no podemos 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 debemos de hacer (o, en general, de usar lenguaje normativo). Llamaré **trascendentes** a enunciados normativos como el ya mencionado “debemos creer cosas verdaderas”, para contrastarlos con enunciados normativos **instrumentales** como, por ejemplo, “debemos creer sólo aquello de lo cual no podemos dudar”. Aun cuando ambos tipos de enunciados nos dicen qué debemos 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 disponemos, mientras que la duda, en tanto carácter subjetivo de nuestro estado mental, es de fácil acceso directo, es decir, nos podemos dar fácilmente 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 1976).



Con esta distinción en mente podemos dar una caracterización más precisa de lo que queremos 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 más 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 veremos más a detalle en la sesión que hablaremos 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 más sobre una situación más 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 trascendente 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 caractericemos un buen argumento dependerá de si queremos una caracterización que nos permita construir nuestros propios argumentos o si lo que queremos es saber cómo evaluar los argumentos de otros (Morado 2000).

Quiero dejar intencionalmente impreciso qué entenderemos 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 dibujaremos 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 **álguien** 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 hacemos lo que hacemos 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, 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 y tras varios tiros su pelota sigue fuera del hoyo. ¿Cuál es su meta relevante? Al no meter la pelota al hoyo, ¿ha alcanzado su meta o no? Uno podría pensar que la meta 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,

pues el objetivo del juego de golf es autónomo e independiente de las intenciones de quien lo juega. 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 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)<sup>2</sup> y, en particular, que tiene entre sus reglas constitutiva el tener como meta el ser conocida (o creída, o verdadera, etc.). Esto significa que si afirmamos algo que no sabemos (o que no creemos, o que no es verdadero, etc.) habremos hecho algo malo, 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-etica 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.<sup>3</sup>

En años recientes, la idea de que hay valores implícitos en ciertos tipos de actos, independientemente de las intenciones de los agentes también ha sido incorporada a ciertos debates en ética aplicada. Rae

---

<sup>2</sup>. “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.”

<sup>3</sup>. Nótese que aquí no estoy haciendo una diferencia entre lo que Steglich-Petersen 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.

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 creo que 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.

## 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* ( $P(x)\text{-implica-}Q(x)$ ) es verdadera o falsa, es necesario y suficiente eliminar la posibilidad de un contraejemplo *a* que sea *P* pero no *Q*. Por ejemplo, si queremos 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* ( $P(x)\text{-implica-}Q(x)$ ) 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 ponemos 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 entendemos 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 más 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 debemos 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 más bien la racionalidad no tiene que ver con reglas lógicas sino con otras cosas que más 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 más 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, podemos 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 podemos 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 entendemos 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 podemos hacer, por ejemplo, pruebas lógicas absolutamente válidas), no siempre podemos darnos ese lujo y en este sentido dichos recursos no deben contarse como realmente disponibles en el sentido relevante.

De esta manera podemos dar cuenta de lo que sucede en el caos cognitivo de una manera que evite ambos cuernos del dilema. Por un lado, reconocemos 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 usar de esta manera estos términos no sea metafórica, 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 en unos casos como en 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 podremos excluir por completo el poder dudar de su verdad.” (Hetherington 2017) 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. Podemos 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. Consideremos 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, podemos 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 podemos saber* cuando o donde habrá un intento de ataque con camiones, no podemos 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 espitémico 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, sabremos 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 (2015) presenté otro ejemplo de relación filosófica importante que es útil pensemos 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 veremos 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.

Si bien la infalibilidad es una virtud extrema de las herramientas, métodos, medios, etc. esto no significa que las herramientas (métodos, medios, etc.) falibles sean inútiles.<sup>4</sup> Es posible que una herramienta sea falibles y aun así sea útil y su existencia valga la pena. Aun cuando las medidas sugeridas a la AFP arriba

---

<sup>4</sup>. 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.

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 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 hablaremos en la siguiente sección.

## F. 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.

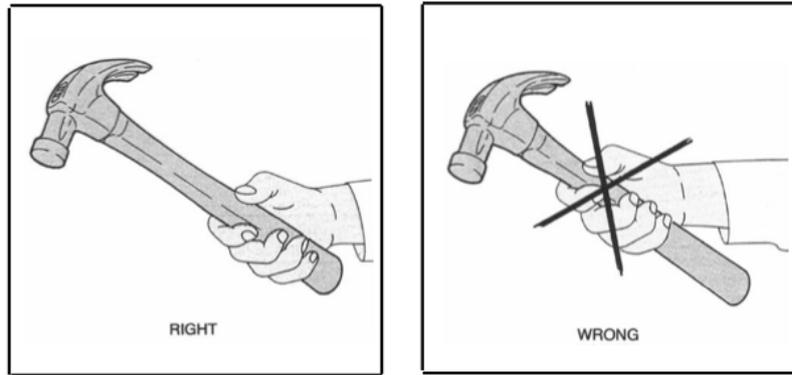


Existe una manera tradicional, y de sentido común, de pensar el lenguaje, entre otras cosas, como un medio de representación, es decir, como una manera de representar cómo es el mundo. Los predicados, por ejemplo, nos sirven para describir maneras en que las cosas pueden ser o no ser. Hay cosas que son 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 o Strawson 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 apliquemos 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 podemos 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 podemos 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. Tomemos un ejemplo sencillo de herramienta: un martillo. Supongamos que alguien quiere enseñarle a otra persona a usar un martillo. Probablemente lo que haga sea demostrarle 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 queremos clavarlo, etc. Sin embargo, imaginemos 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 clavos 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 aprendemos a usar el martillo, como casi cualquier herramienta, no lo hacemos a través de reglas exactas, sino a través de una combinación de demostración 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 es 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. 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 identificar un objeto, algo habra salido mal si no logro que mi escucha identifique al objeto al 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 las siguientes sesiones veremos 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 conocemos como los “casos Gettier” en honor al filosofo que los presentó de manera elegante y sugerente en un breve artículo de título “¿Es el conocimiento creencia verdadera justificada?”, el segundo es el fenómeno de la suerte moral. Aún cuando, en realidad, ambos fenómenos no son sino casos particulares de un fenómeno más general, nos acercaremos 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 guíarnos 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 sólo 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 actuemos puede tener consecuencias de gravedad tal que justifiquen le dediquemos cuidado a nuestras decisiones. Existe la normatividad epistémica, por ejemplo, porque no es indiferente qué pensemos, es decir, qué creamos, qué inferamos, qué dudemos, 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 sobrevivencia como individuos, sino también como especies, por ejemplo. Es por eso que nos preguntamos qué debemos de hacer, como debemos 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 usásemos 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 podemos 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 o herramienta es adecuado si, a partir de recursos que está 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, al punto de que sea confiable esperar alcanzar nuestros fines.

Ahora bien, una vez que reconocemos 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 obtenemos nuestro fin.
2. No usamos un medio adecuado y no obtenemos nuestro fin.
3. Usamos un medio adecuado y, sin embargo, no obtenemos nuestro fin.
4. No usamos un medio adecuado y, aun así, obtenemos 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 fablista comúnmente no dicen nada sobre (4). En esta sección del curso veremos que esta división de cuatro tipos de situaciones no es suficientemente fina para dar cuenta de nuestras verdaderas prácticas normativas. Veremos que no todos los casos de tipo (1), por ejemplo, son normativamente equivalentes, por ejemplo, veremos que no siempre que se da (1) podemos 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);<sup>5</sup> 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 *apriori* 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.

---

<sup>5</sup>. 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

- 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).

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 debemos 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. Llamemos 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  $(pvq)$  o  $((pvq) vr)$ , 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 – o que fuera infalible. Sin embargo, ya hemos 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.

Sin embargo, 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 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 ganemos 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.<sup>6</sup> 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 juicio (de que somos inocentes, pues actuamos de manera cuidadosa y responsable). Cuando un acto es in-impugnable según el primer criterio, pero no según el segundo, tenemos un desafortunado caso de *mala suerte*.

En general, hay dos maneras de concebir la satisfacción del primero criterio normativo. Desde la

---

<sup>6</sup>. 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).

perspectiva internista, 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 externista, una persona está justificada en tomar un riesgo si el cálculo de costos y beneficios, dadas las probabilidades de obtención de cada resultado 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 externista de lo que hace tomar un riesgo racional o no capture nuestra intuiciones acerca de lo que es racional esperar de un agente cuidadoso. Algunos filósofos han querido argumentar que un tipo de riesgo es más básico que el otro, o que puede reducirse a él; la cuestión no me interesa y sólo quería señalar que, de hecho, existen ambos enfoques en la literatura.

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. 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 hemos aprendido algo de veinticinco siglos de escepticismo filosófico en occidente es que no podemos 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, hemos satisfecho los requisitos que impone el primer criterio de normatividad (hemos tratado de reducir el riesgo de falsedad, cumplido nuestras obligaciones epistémicas, etc.). Si nuestra creencia es verdadera, hemos 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.

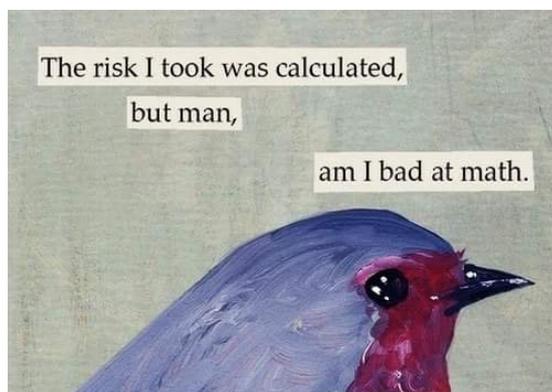
Poner la distinción en términos epistémicos nos sirve además para recordarnos que existe un tercer criterio normativo involucrado en nuestra evaluación de actos arriesgados. Si bien es cierto que hay algo malo en tomar riesgos injustificados, o que algo malo sucede si no obtenemos el resultado deseado 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, argumento que ignorar esta distinción es un tipo de error detrás del relativismo de John MacFarlane, 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, podremos 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 podremos acusar a dichas fuerzas de negligentes, sino que lo más adecuado sería decir que se tuvo mala suerte.



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, podemos 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 proponemos algo como objetivo o emprendemos una tarea, no basta tener éxito y conseguir eso que nos proponemos 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 ahorrando, 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 *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), 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, 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. Regresemos 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. Adoptemos la convención de llamar “consecuencia” a la relación que hay entre un esfuerzo virtuoso por lograr algo y el logro que se logra de él. 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 Hilary Kornblith llama el *consecuencialismo burdo*, 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 independiente del valor de aquello que logran. En otras palabras, el esfuerzo que no termina en éxito no vale nada. Si valoramos la posesión o el ejercicio de una capacidad no es sino porque valoramos obtener aquello que podemos alcanzar con el ejercicio de dicha capacidad. Por lo tanto, para el *consecuencialismo*, no debemos valorar diferente un logro de un accidente afortunado. En el otro extremo se encuentran filósofos de inclinación *Kantiana* (pienso en Christine Korsgaard, por ejemplo) para quienes la diferencia fundamental entre el logro y el accidente afortunado es que sólo el primero puede adjudicarse normativamente al sujeto. Por eso no elogiamos a quienes ganan la lotería (aunque 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).

En contraste, el trabajo de John Grecco, Peter Graham, Ernest Sosa, etc. se dirige, mas bien, a responder a la pregunta descriptiva. La cuestión es especialmente complicada porque al tratar de definir un logro como, por ejemplo, aquel que se debe al ejercicio de una capacidad, el problema pasa a determinar (de una manera no circular) qué es una capacidad 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. 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 se 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 insignificante, y ésta es una noción irreduciblemente normativa, no sólo descriptiva. 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 fallé, 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 cuando se vive en zona de huracanes es racional 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 podemos garantizar que, si se da dicha situación, obtendremos el resultado (si así lo deseamos), pero esto no significa que si se da una situación para la que no estamos preparados, podemos garantizar que **no** se va a obtener el resultado, sino que **no** podemos 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.

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 realizo 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.

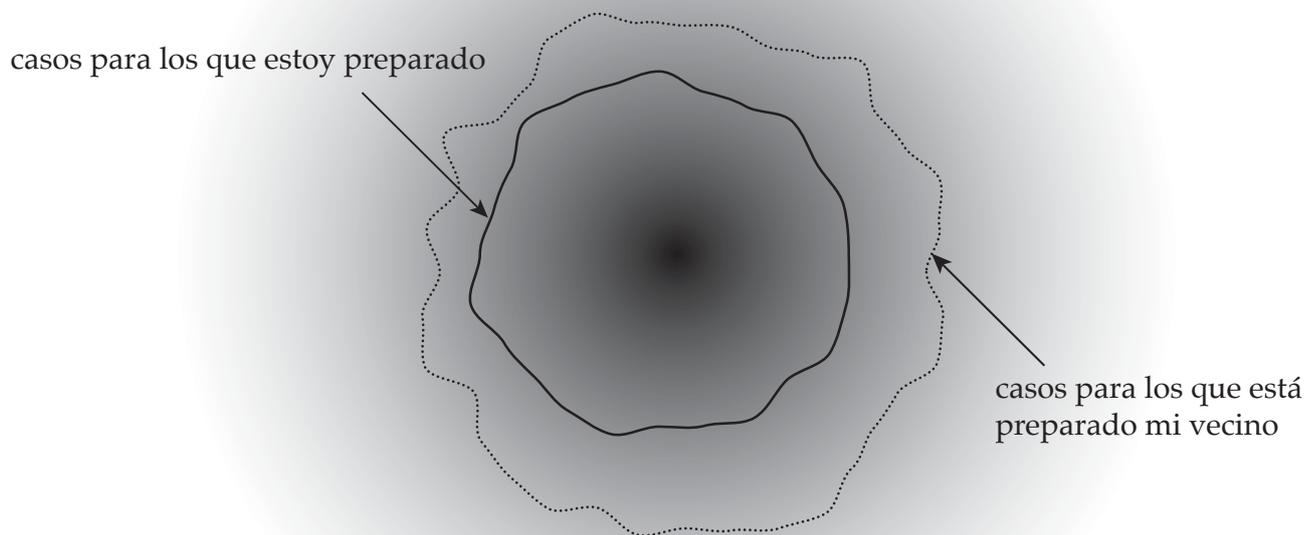
Pongamos un ejemplo para ilustrar esto: Supongamos que vivo en una zona de tormentas severas, donde es un riesgo presente la posibilidad de inundaciones. Si soy una persona responsable, contactaré a las autoridades adecuadas para enterarme si mi vivienda se encuentra en una zona de derrumbes o inundación 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.

Podemos usar el siguiente diagrama para explicar mejor esta idea. Imagínense este plano como el espacio de posibilidades de tal manera que cada punto de él representa una situación futura relevante ordenada de tal manera que entre más cercana al centro, más previsible es. 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 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 hemos 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 de mi vecino es mas 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, estemos 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.

Cambiemos un poco el ejemplo e imaginemos 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 debemos decir que

estábamos preparados para la inundación y que la sobrevivimos, pero es claro que no podemos 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 hacemos, no será porque estábamos preparados o porque no lo estábamos, sino por suerte: buena suerte si sobrevivimos, mala suerte si no lo hacemos.

Traslademos 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,<sup>7</sup> pero no podemos 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.<sup>8</sup> 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 podremos 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

---

7. 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.

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

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. 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 creemos, por ejemplo, que es bueno estar delgado, por ejemplo, podemos distinguir entre casos en los que se es afortunado ser delgado de aquellos en los que es un logro; si creemos que la sopa que estamos comiendo es deliciosa, podemos 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, podemos distinguir casos en los que “hasta el mejor 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 podemos generar fácilmente casos de Gettier. Podemos 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	CONOCIMIIENTO
ACCIÓN DIRIGIDA HACIA UN OBJETIVO	ALCANZAR EL OBJETIVO	Haber hecho lo suficiente para lograr el objetivo	LOGRO
EMPRESA	ÉXITO	PREPARACIÓN	
ACCIÓN MORALMENTE EVALUABLE	BUENO EN SUS CONSECUENCIAS	RESPONSABILIDAD (NO NEGLIGENCIA)	
Aserción	verdad	justificación	

Aunque todos estos son espacios en los cuales podemos 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 le interesa la ciencia y la verdad como criterio para evaluarla. A la filosofía del lenguaje, en contraste, le interesan los actos de aserción y, otra vez, la verdad de toque se dijo como criterio

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 podemos aplicarla a otros campos del lenguaje de interés filosófico. Por ejemplo, como ya mencionaba en la sección anterior, también podemos 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 estemos preparados para hacerlo en los casos más previsible. Es por ello que nuestros conceptos son vagos. Así pues, podemos 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.

### C. Hay casos Gettier en Matemáticas?

Una vez que hemos entendido mejor las moralejas extraíbles del falibilismo y el 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. ¿Podemos 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.

Para empezar esta reflexión, vale la pena empezar preguntándose si la noción de “prueba” relevante para la epistemología de las matemáticas es una noción trascendente instrumental, es decir, si la regla de que debemos 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 caracterizable en terminos 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 dos requisitos que buscaba Hilbert, es decir, que fuera infalible e instrumental. Esto significa que la noción de prueba operativa en matemáticas es una noción trascendente o falible.

Asumamos que la noción de ‘prueba’ operativa en matemáticas es trascendente. Esto significa que 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 hemos probado un teorema, sino sólo que lo hemos 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 exámen una secuencia de fórmulas al azar las cuales, afortunadamente para él, terminasn 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. Podemos 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 oír 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. Para fundamentar esta concepción de la prueba matemática, revisemos 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 podemos 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 otro problema sustancial 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.

## REFERENCIAS

- Alston, J. (2000), *Illocutionary Acts and Sentence Meaning*. Ithaca: Cornell University Press.
- 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 (2017) "Inferencia y Presuposición", manuscrito bajo dictamen.
- Barceló, Axel y Robert Stainton (2018) "Quasi-Factives and Cognitive Efficiency", manuscrito bajo dictamen.
- Brandom, Robert (1994), *Making it Explicit*. Cambridge: Harvard University Press.
- 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.
- Douven, I. (2006), "Assertion, knowledge, and rational credibility", *Philosophical Review* 115: 449–485.
- 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.

- Eraña, Ángeles (2003) “Normatividad epistémica y estructura heurística del razonamiento”, *Crítica, Revista Hispanoamericana de Filosofía*. Vol. 35, No. 104 (agosto 2003): 69–108
- 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
- Greco, John (2010) *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*, Cambridge University Press.
- Hansson, Ove; Karin Edvarsson Björnberg y John Cantwell (2016) “Self-Defeating Goals”, *Dialéctica*, Volume 70, Issue 4, December 2016, Pages 491–512
- Hetherington, Stephen (2017) “Fallibilism”, *The Internet Encyclopedia of Philosophy*, ISSN 2161-0002, <http://www.iep.utm.edu/>, consultada el 7 de enero.
- 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.
- Langton, Rae & Caroline West (1999) “Scorekeeping in a pornographic language game”, *Australasian Journal of Philosophy*, 77:3, 303-319
- Martin, Justin W. & Fiery Cushman (2015) “The Adaptive Logic of Moral Luck”, en Justin Sytsma y Joshua Knobe (eds.) *Blackwell Companion to Experimental Philosophy*.
- 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
- Nozick, R. 1993, *The Nature of Rationality*, Princeton, NJ: Princeton University Press.
- 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.

- Rescorla, Michael (2009), "Assertion and its Constitutive Norms", *Philosophy and Phenomenological Research*, Vol. LXXIX No. 1, July 2009, pp. 98 - 130.
- Rudy, Fernando (manuscrito)
- Ryle, Gilbert (1953) "Ordinary Language", *The Philosophical Review*, Vol. 62, No. 2 (Apr., 1953), pp. 167-186
- 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.
- 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.
- Velleman, D. (2000) *The Possibility of Practical Reason*, Oxford University Press.
- Williamson, Timothy (2000) *Knowledge and its Limits*. Oxford: Oxford University Press.
- Robin Zheng y Nils-Hennes Stear (manuscrito) "Imagining in oppressive contexts, or, what's (intrinsically) wrong with "Blacking Up"?"