Rigorismo e Integridad

Un ensayo sobre la concepción del yo dividido en la filosofía práctica kantiana

José Leonardo González Valderrama

Universidad Nacional de Colombia
Facultad de Ciencias Humanas, Departamento de Filosofía
Bogotá, Colombia
2020

Rigorismo e Integridad

Un ensayo sobre la concepción del yo dividido en la filosofía práctica kantiana

José Leonardo González Valderrama

Tesis presentada como requisito parcial para optar al título de:

Doctorado en Filosofía

Director (a):

Ph.D. Ángela Uribe Botero

Línea de Investigación:

Ética

Grupo de Investigación:

Relativismo y racionalidad

Universidad Nacional de Colombia

Facultad de Ciencias Humanas, Departamento de Filosofía

Bogotá, Colombia

2020

Resumen

La filosofía moral kantiana se ha asociado usualmente al rigorismo moral. Según esta posición, los deberes morales deben ser cumplidos incondicionalmente. Por ejemplo, afirma Kant, debemos cumplir el deber moral de la veracidad incluso ante un asesino que nos pregunte por el paradero de alguien que se refugia en nuestra casa. El objetivo de esta investigación es realizar una crítica al rigorismo moral kantiano. Para ello he dividido el texto en tres partes. En la primera parte presento las posición de Kant sobre la mentira (capítulo 1) y tras analizar el procedimiento del imperativo categórico explico en qué sentido este autor está comprometido con el rigorismo moral (capítulo 2). En la segunda parte, muestro que el rigorismo moral está fundamentado en el modelo de la agencia del yo dividido. Según este modelo, la agencia es un ejercicio de control por parte de un yo racional que debe controlar o someter a la parte pasiva no racional (capítulo 3). Siguiendo el trabajo de Bernard Williams, muestro que este modelo agencial conduce al problema de la integridad personal, pues en ocasiones la moralidad nos exige actuar en contra de nuestro carácter y de aquello que configura nuestra integridad (capítulo 4). En la tercera parte, examino la respuesta de Christine Korsgaard al problema de la integridad. Esta autora rechaza la concepción del yo dividido, y en cambio propone el modelo constitucional. Según este modelo, ser agente solo es posible cuando este está constituido como un todo integrado, y solo es posible estar integrado si existe un compromiso con la moralidad. De esta manera, ella puede articular una respuesta al problema de la integridad sin abandonar lo que denomino "la tesis de la prioridad normativa de la moralidad" (capítulo 5). En contra de esta posición, y siguiendo a Josep Corbí, articulo una crítica al modelo constitucional mostrando que este todavía está comprometido con la concepción del yo dividido (capítulo 6.1). Corbí también realiza una crítica a la tesis de la prioridad normativa de la moralidad (6.2 y 6.3). Al final concluyo que la filosofía moral kantiana, ya sea en su versión clásica o en la versión contemporánea de Korsgaard, no puede responder al problema de la integridad personal, en la medida en que desestima injustificadamente el valor normativo de fenómenos prácticos no morales.

Palabras clave: integridad personal, rigorismo moral, Kant, Christine Korsgaard, Bernard Williams

Abstract

Kantian moral philosophy has usually been associated with moral rigorism. According to this position, moral duties must be fulfilled unconditionally. Kant claims that we must fulfill the moral duty of truthfulness even in the face of a murderer who asks us about the whereabouts of someone who takes refuge in our house. The aim of this investigation is to make a criticism of Kantian moral rigorism. To this purpose, I have divided the text in three parts. In the first part, I present Kant's position on lying (chapter 1), and after analyzing the procedure of categorical imperative I explain in which sense this author is committed to moral rigorism (chapter 2). In the second part, I show that moral rigorism is based on the conception of the divided self. According to this model, agency is an exercise of control by a rational self that must control the non-rational passive part (chapter 3). Following to Bernard Williams, I show that this model leads to the problem of personal integrity, since sometimes morality requires us to act against our character and against that which configures our integrity (chapter 4). In the third part, I examine Christine Korsgaard's response to the problem of integrity. She rejects the model of the divided self and proposes instead the constitutional model. According to this model, being an agent is only possible when it is constituted as an integrated whole, and it is only possible to be integrated if there is a commitment to morality. In this way, she can articulate a response to the problem of integrity without abandoning what I call "the thesis of the normative priority of morality" (chapter 5). Contrary to this position, and following Josep Corbí, I show that the constitutional model is still committed to the conception of the divided self (chapter 6.1). In addition, I also make a critique of the thesis of the normative priority of morality (6.2 and 6.3). I conclude that Kantian moral philosophy, whether in its classical version or in Korsgaard's contemporary version, cannot respond to the problem of personal integrity because it unjustifiably disregards the normative value of practical non-moral phenomena.

Keywords: personal integrity, moral rigorismo, Kant, Christine Korsgaard, Bernard Williams

Tabla de contenido

INTRODUCCIÓN **10**

I. DELIMITACIÓN CONCEPTUAL: KANT Y LA MENTIRA **16**

1. Kant: sobre la mentira **17**

1.1 *Lecciones de ética* 18

1.1.1. La mentira en caso de necesidad 19

1.2 "Sobre el supuesto derecho a mentir por amor a la humanidad" 22

1.2.1 El deber de ser veraz ante el asesino en la puerta 23

1.2.2 Contexto histórico-político de "Sobre un supuesto derecho a mentir" 27

1.3 *Metafísica de las costumbres* 31

1.3.1 Deberes hacia uno mismo 32

1.3.2 La mentira como violación del deber hacia uno mismo 33

1.3.3 ¿Qué hacer frente al asesino en la puerta? 37

2. Deberes perfectos y rigorismo kantiano **40**

2.1 Acción por deber y acción conforme al deber 41

2.1.1 Buena voluntad y deber 41

2.1.2 Acción por deber y acción conforme al deber 44

2.1.3 Sobredeterminación motivacional [*motivational overdetermination*]: 50

2.1.4 Imperativo como deber-ser 52

2.2 Imperativo categórico 53

2.2.1 El procedimiento del imperativo categórico 56

2.3 Deberes perfectos y deberes imperfectos 64

2.4 La máxima de la mentira al asesino 67

2.5 La mentira: usar al otro como medio 72

2.6 Korsgaard: el derecho a mentir 76

II. RIGORISMO MORAL KANTIANO Y EL PROBLEMA DE LA INTEGRIDAD PERSONAL **83**

3. Rigorismo kantiano 84

3.1 El rigorismo moral kantiano 85

3.2 Develando el rigorismo 89

3.2.1 El requisito prescriptivista 90

3.2.2 La imagen desencantada del mundo 92

3.3 El perspectivismo kantiano 98

3.4 La concepción del yo dividido 101

3.4.1 El concepto kantiano de libertad en la *Fundamentación de la metafísica de las costumbres*. 102

3.4.2 La noción de libertad en la *Metafísica de las Costumbres*: ¿cómo imputar acciones no morales? 104

3.4.3 La concepción del yo dividido 109

4. El problema del rigorismo moral 116

4.1 Williams: el problema de la integridad personal 116

4.2 Integridad: carácter y autoindulgencia 120

4.2.1 Integridad: identidad y carácter 121

a. Identidad y carácter 123

b. 'Seguir hacia adelante' y deseos categóricos 125

c. Proyectos fundamentales y sentido [*meaning*] 127

d. Amenaza a la integridad: proyectos fundamentales y carácter 133

e. Muerte práctica 136

4.2.2 Autoindulgencia moral 137

4.3 El problema de la integridad y la concepción del yo dividido 140

III. ¿SUPERA EL MODELO KANTIANO EL PROBLEMA DE LA INTEGRIDAD? 145

5. Christine Korsgaard: integridad como auto-constitución 146

5.1 Agencia e integridad 151

5.2 Modelo constitucional de la agencia 154

5.2.1 Modelo constitucional de la agencia 154

5.2.2 El modelo constitucional y el agente kantiano 160

5.2.3 Identidades prácticas 164

a. Contingencia 166
b. Identidad práctica humana 168
c. Identidad moral 171
5.2.4 Razones públicas y razones privadas 172
a. La naturaleza pública de las razones 176
b. Ley moral 178
c. Integridad y unidad 181
5.3 Integridad como auto-constitución 183

6. Crítica a la primacía normativa de la moralidad: sentido y necesidad práctica 187
6.1 Crítica al modelo constitucional 190
6.1.1 Identidades prácticas y asimetrías 191
6.1.2 La tesis de la indispensabilidad de los principios universales 195
6.2 ¿Primacía normativa de la moralidad? 199
6.2.1 Prioridad constitutiva de la moralidad 200
a. Schmagencia [*Schmagency*] y el juego de la agencia 202
b. 'Razón para estar vivo' y moralidad 206
c. ¿Primacía constitutiva de la moralidad? 214
d. Reivindicación de la crítica de Williams: moralidad y sentido 215
6.2.2. Tesis de la prevalencia 217
a. La distinción entre ética y moralidad 219
b. El concepto de importancia 224
c. Prioridad deliberativa 229
d. Necesidad práctica 231
e. Obligación moral y no-moral 240

Conclusiones 245

Bibliografía 252

Rigorismo e Integridad

Un ensayo sobre la concepción del yo dividido en la filosofía práctica kantiana

Quino

"El que obedece no se oye a sí mismo"

Nietzsche, *Así habló Zarathustra*

Método de citación de las obras de Kant

Para las referencias a las obras de Kant adoptaré las abreviaciones aquí indicadas seguidas de la paginación de la Academia Alemana de Ciencias (*Kants gesammelte Schriften*).

A	*Antropología en sentido pragmático* (*Anthropologie in pragmatischer Hinsicht abgessat*, 1789)
C	*Correspondencia* (*Kant's Briefwechsel*, 1749-1800)
C1	*Crítica de la razón pura* (*Kritik der reinen Vernunft*, 1ª ed., 1781; 2ª ed., 1787)
C2	*Crítica de la razón práctica* (*Kritik der praktischen Vernunft*, 1ª ed., 1781; 2ª ed., 1788)
F	*Fundamentación para una metafísica de las costumbres* (*Grundlegung zur Metaphysik der Sitten*, 1785)
LE	*Lecciones de ética* (*Eine Vorlesung Kants über Ethik*, 1775-1781)
MC	*Metafísica de las costumbres* (*Die Metaphysik der Sitten*, 1797)
PP	*Sobre la paz perpetua* (Zum ewigen Frieden, 1795)
R	*La religión dentro de los límites de la mera razón* (*Die Religion innerhalb der Grenzen der blossen Vernunft*, 1793)
SDMF	"Sobre un presunto derecho a mentir por filantropía" (*Über ein vermeintes Recht, aus Menschenliebe zu lügen*, 1797)

Introducción

Tengo la sospecha de que los seres humanos le damos una excesiva importancia a la moral en nuestras vidas. No dudo que la moral sea importante y fundamental para nuestra vida práctica. Actuamos, juzgamos y vivimos de acuerdo con ella. Mi sospecha, sin embargo, es que nuestra comprensión de la moralidad ha sobre-valorado su papel en nuestras vidas, y esto, creo, distorsiona la comprensión de nosotros mismos y de nuestras acciones.

La filosofía moral de Kant es una confirmación de esta sospecha. Esta concepción de la moralidad, especialmente la que se encuentra expuesta en la *Fundamentación de la metafísica de las costumbres*, representa, en mi interpretación, la idea de que la moral es el valor más importante de la vida práctica de los seres humanos.

Si mi sospecha sobre la filosofía moral de Kant es correcta, me parece importante dar cuenta de las razones por las que considero que ella constituye una moralización excesiva de la vida humana. De ser esto cierto, también me parece importante establecerle un *límite* a la moralización. ¿Acaso la vida práctica de los seres humanos se agota en lo que estamos moralmente obligados a hacer? ¿Son las normas y deberes morales las únicas guías normativas en nuestra vida? Mi intuición, que espero defender en esta investigación, es que ni la vida práctica humana se agota en la moralidad, ni los deberes morales son lo único a lo que estamos obligados.

En esta investigación me propongo llevar a cabo una crítica a la filosofía práctica kantiana. La tarea, así descrita, es titánica. Por lo tanto, he decidido centrar el trabajo en un objetivo más modesto y preciso: criticar el rigorismo moral kantiano. El rigorismo es la tesis moral que afirma que los deberes morales deben ser cumplidos incondicionalmente. Esta tal vez sea una de las interpretaciones más problemáticas y discutibles de la filosofía práctica de Kant. Una primera dificultad para cumplir este objetivo es que no es evidente que la filosofía moral kantiana pueda leerse en términos rigoristas. Aun así, esta intepretación se ha extendido lo suficiente como para

que, incluso, varios autores kantianos se hayan esforzado por desestimarla[1]. Y concuerdo con ellos en que no se puede dar por sentada esta interpretación, y en que, de ser correcta, ella requiere de varios matices, aclaraciones y precisiones. En todo caso, y a diferencia de estos autores kantianos, me comprometo y defiendo la lectura rigorista del pensamiento kantiano.

La lectura rigorista, más allá de si es correcta o no, es una articulación extrema, radical, de la filosofía práctica de Kant. La ventaja metodológica de esta radicalidad es que somete al sistema kantiano a variables y condiciones extremas para observar cómo se comporta y así descubrir sus límites y posibilidades. En este caso, como si de un experimento se tratara, intento llevar la filosofía moral kantiana hasta sus límites, y es ahí, en el extremo de sus posibilidades, donde se descubren las aristas tan afiladas de esta teoría y se entrevén algunas de sus fisuras.

El malestar que provoca el rigorismo en la escena académica se debe, en parte, a que choca directamente contra nuestras intuiciones de orden práctico. Este choque no necesariamente es un problema. Cuando un análisis filosófico va en contra de nuestras intuiciones, siempre es posible insistir en que estas son incorrectas. Por esta razón, por *sí mismas*, las intuiciones no pueden ser una razón válida para desvirtuar un sistema filosófico. Sin embargo, a pesar de su falta de valor argumentativo, las intuiciones mismas no son inútiles, pues orientan e impulsan la crítica, y lo hacen aun más cuando salen a la luz con tanta fuerza, como en el caso del rigorismo. A mi modo de ver, el choque que genera el rigorismo con nuestras intuiciones es un indicio de que la propuesta kantiana falla. El reto está en caracterizar y justificar filosóficamente esta falla.

La forma en la que desarrollo filosóficamente esta intuición es mostrando que el rigorismo moral kantiano, al imponer demandas morales incondicionales sobre los agentes, los obliga a subordinar cualquier otro criterio práctico a las exigencias morales. Me concentro específicamente en el conflicto que puede darse entre el rigorismo moral y la integridad personal: las exigencias incondicionales de los deberes perfectos pueden constituir una amenaza a la integridad de una persona. Considero que el rigorismo kantiano es una de las versiones filosóficas más radicales de la moralización de la vida humana, a tal punto que otro tipo de fuentes normativas se ven amenazadas o excluidas. Si consigo demostrar la importancia y el valor normativo de la integridad,

[1] Ver especialmente: Paton (1954), Korsgaard (2000b), Shapiro (2006), Cholbi (2009; 2013).

entonces estaría estableciendo un límite a la moralidad kantiana, estaría demostrando que esta concepción de la moralidad no es el único modelo normativo para guiar nuestras acciones.

Para llevar a cabo mi objetivo de criticar el rigorismo moral kantiano, he dividido la investigación en tres partes, cada una de ellas, compuesta por dos capítulos. En la primera parte, justifico por qué la filosofía moral kantiana está comprometida con el rigorismo moral. En el primer capítulo, examino la posición de Kant con respecto a la mentira, específicamente me centro en el famoso caso del asesino en la puerta que se discute en el texto "Sobre el presunto derecho a mentir por filantropía". Para Kant, si un asesino nos pregunta por el paradero de un amigo que se refugia en nuestra casa, tenemos el deber de decirle la verdad (§1.2). Este caso es relevante porque usualmente ha servido a los detractores de Kant como un recurso para justificar la acusación de rigorismo que pesa en contra de la filosofía kantiana. Sin embargo, muestro que la última palabra de Kant sobre la mentira no está en este texto. En realidad hay otras fuentes bibliográficas que llegan a desvirtuar la acusación de ser la filosofía moral de Kant rigorista (§1.1 y §1.3). En el segundo capítulo, en contraste, justifico la lectura rigorista de la filosofía kantiana. Mi estrategia consiste en examinar el procedimiento del imperativo categórico (§2.1 y §2.2), y muestro que el rigorismo moral se deriva de la noción de deberes perfectos (§2.3 y §2.4). En este capítulo, además, respondo a la crítica que hace Christine Korsgaard en contra de la lectura rigorista de la filosofía práctica de Kant (§2.5 y §2.6).

En la segunda parte de la investigación, planteo el problema del rigorismo moral kantiano: sostengo, apoyado en el pensamiento de Bernard Williams, que el rigorismo representa una amenaza a la integridad personal. El tercer capítulo está dedicado a definir y a caracterizar con precisión el rigorismo (§3.1). En este capítulo muestro, además, que el rigorismo está soportado en el modelo que concibe al yo como dividido. Una de las intuiciones que desarrollo en esta segunda parte de la investigación es la idea de que existe en Kant una estrecha relación entre la metafísica, la antropología y la psicología moral (§3.2). En el caso específico de la filosofía práctica kantiana, resulta que el idealismo trascendental kantiano conduce a un modelo de la agencia en el que el yo está dividido, y a su vez esto configura la psicología moral de un agente que debe cumplir incondicionalmente con el deber moral (§3.3). Esta estrecha relación responde a lo que Bernard Williams denomina la "psicología metafísica del moralismo kantiano" [*Metaphysical phsycology of*

kantian moralism] (Pyle, 1994, p. 12). Esto, a su vez, hace posible el tránsito entre distintas categorías: por un lado, muestro que a partir de la tesis moral del rigorismo es posible derivar el modelo agencial del yo dividido; y por otro lado, del rigorismo también se desprende una psicología moral en el que el agente, a pesar de experimentar el conflicto entre su razón y su inclinación, está constreñido por las exigencias incondicionales del deber moral (§3.4). Busco, entonces, mostrar que el rigorismo moral kantiano, en virtud de la concepción del yo dividido que lo soporta y en virtud del tipo de psicología moral que ese yo implica, representa una amenaza a la integridad personal.

En el cuarto capítulo, me detengo a explicar qué es la integridad personal y en qué sentido el rigorismo kantiano representa una amenaza contra ella (§4.1). Siguiendo a Bernard Williams, caracterizo la integridad personal en términos del carácter personal. El concepto de carácter lo relaciono con los proyectos fundamentales que tenemos los seres humanos y, en particular, a través de las razones que solemos tener las personas para permanecer en la vida (§4.2). Mi tesis es que el rigorismo moral, al exigir el cumplimiento incondicional del deber moral, le pide al agente ir en contra de su integridad, lo cual compromete negativamente su identidad, su carácter, sus proyectos fundamentales y las razones que tiene para vivir (§4.3).

En la tercera parte de la investigación traslado la discusión del rigorismo a un escenario más contemporáneo. La versión "clásica" del rigorismo tiene su forma más acabada en Kant. No obstante, en la escena contemporánea, encuentro que el trabajo de Christine Korsgaard (2009) constituye una versión matizada y más sofisticada –que la del propio Kant– del rigorismo kantiano. En el capítulo 5, presento la respuesta que Christine Korsgaard ofrece al problema de la integridad personal. Esta autora se distancia del modelo del yo dividido porque lo considera inadecuado para explicar la agencia kantiana (§5.1). En su reemplazo, postula un modelo de la agencia alternativo al que llama *constitucional* (§5.2). Según este modelo, un agente es quien actúa como un 'todo integrado', con lo cual Korsgaard articula una noción de agencia en términos de integridad, e interpreta la integridad como unidad. Adicionalmente, esta autora postula que aquello que unifica o integra al agente son los principios morales, por eso, concluye que estos principios son constitutivos de la agencia misma: no es posible ser agente a menos que una persona esté comprometida constitutivamente con los principios morales (§5.3).

El capítulo 6 tiene como objetivo responder críticamente a los planteamientos de Korsgaard. Esta crítica se realiza en dos momentos. Siguiendo a Josep Corbí, en el primer momento muestro que el modelo constitutivo ofrecido por Korsgaard en realidad sigue comprometido con una concepción del yo dividido. En este sentido, cuestiono la posibilidad de que Korsgaard pueda articular la noción integridad en términos de unidad (§6.1). En el segundo momento, muestro que Korsgaard todavía comparte con el rigorismo kantiano lo que denomino la *tesis de la primacía normativa de la moralidad* (PNM). Esta tesis sostiene que la moralidad tiene prioridad normativa por encima de cualquier otro criterio práctico (§6.2). Mi objetivo es demostrar que PNM es falsa, y para ello distingo entre una versión fuerte y otra débil de esta tesis. Korsgaard defiende la versión fuerte a través de su compromiso con lo que denomino *la primacía constitutiva de la moralidad*; es decir, la tesis según la cual la normatividad de la moralidad es absoluta e incondicionada en virtud de que ella es constitutiva de la agencia. Mi crítica va encaminada a mostrar que la incondicionalidad de la moralidad es falsa porque el compromiso de un agente con la moralidad estaría condicionado a que él tenga una razón para estar vivo (§6.2.1). La versión débil de PNM es la *tesis de la prevalencia*. Esta tesis admite la posibilidad de que la moralidad no sea incondicionada, pero mantiene la idea de que sí es el criterio normativo superior o más importante en la deliberación práctica. Mis argumentos contra esta tesis pasan por trazar la distinción que hace Williams entre ética y moralidad, y por demostrar que la moralidad kantiana reduce las consideraciones éticas a la noción de deber moral. En contra de lo que llamo el *reduccionismo deontológico*, introduzco los conceptos normativos éticos de importancia y prioridad deliberativa, para, con ellos, articular la noción de necesidad práctica. Quisiera, con esto, hacer evidente que los seres humanos, en ocasiones, nos vemos impelidos a realizar actos por necesidad práctica que, sin embargo, no obedecen a la normatividad impuesta por la moral (§6.2.2). Con esta idea busco concluir que la tesis de la prioridad normativa de la moralidad —que es el núcleo del rigorismo kantiano— es falsa.

El alcance de la crítica al rigorismo moral kantiano constituye una crítica a la comprensión sobrevalorada de la moralidad. Y con esto no busco sugerir que la moralidad debe estar entonces subordinada a criterios éticos, como el de la integridad personal. No pretendo sustituir el rigorismo moral por un rigorismo ético. Mi intención es mucho más simple: no es obvio que la moralidad tenga prioridad por encima de valores o consideraciones no morales. Dicho de otra manera, creo

que es un error pensar que ante la pregunta '¿qué debemos hacer?' la respuesta sea que siempre debemos actuar moralmente.[2]

[2] Aclaro que he traducido todas las citas de los textos originales en inglés. He tomado esta decisión, primero, para acercar a mi lengua materna las ideas de los autores citados y, segundo, para evitarle al lector un cambio de registro idiomático a la hora de abordar mi texto. Para identificar si la fuente original citada está en inglés, remito al lector a la sección Bibliografía donde encontrará todos los datos bibliográficos de la fuente de acuerdo con el sistema de citación APA 7ª edición.

I. Delimitación conceptual: Kant y la mentira

1. Kant: sobre la mentira

1.1 Lecciones de ética

1.1.1 La mentira en caso de necesidad

1.2 "Sobre el supuesto derecho a mentir por amor a la humanidad"

1.2.1 El deber de ser veraz ante el asesino en la puerta

1.2.2 Contexto histórico-político de "Sobre un supuesto derecho a mentir"

1.3 Metafísica de las costumbres

1.3.1 Deberes hacia uno mismo

1.3.2 La mentira como violación del deber hacia uno mismo

1.3.3 ¿Qué hacer frente al asesino en la puerta?

A la filosofía moral kantiana le es atribuido el rigorismo moral. El rigorismo moral, de manera general, puede entenderse como la exigencia del cumplimiento incondicional de los deberes morales. El problema del rigorismo, al menos de manera intuitiva, es que una persona puede llegar a realizar actos que, a la luz del deber, son morales, pero afectan negativamene a otros. Si lo más importante es cumplir con el deber, es posible que se den situaciones en las que el cumplimiento incondicional del deber traiga consecuencias negativas o perjudiciales para otras personas o incluso para quien actúa.

La acusación de rigorismo que se le imputa a la filosofía práctica de Kant se justifica principalmente en la posición de este autor con respecto a la mentira. Para Kant, la veracidad es un deber moral incondicional. Su posición es tan radical que incluso, dice, debemos ser veraces con el asesino que llega a nuestra casa preguntando por el paradero de nuestro amigo que se encuentra refugiado en ella, sin importar las consecuencias incluso mortales que pueda acerrear la veracidad.

Dado que la acusación rigorista se justifica principalmente en la prohibición moral de la mentira, en este primer capítulo me detengo a recostruir y a examinar la posición de Kant con respecto a la mentira. Esta reconstrucción servirá de trasfondo conceptual para la discusión que desarrollaré en el segundo capítulo, donde examinaré hasta qué punto la filosofía moral kantiana está

comprometida con el rigorismo moral o si acaso hay aún espacio para hacer excepciones a los deberes morales. Por lo pronto, la reconstrucción que sigue a continuación estará guiada por un eje cronológico, así que iniciaré con el texto más temprano del autor e iré avanzando hacia los más tardíos.

1.1 *Lecciones de ética*[3]

[3] Las *Lecciones de ética* es un libro producto de la recopilación de las notas de clase de los estudiantes de Kant durante 1777-1794.[a] En sentido estricto, esta obra no fue redactada por Kant, y tal vez por eso ha sido poco atendida. Otra razón para su escasa recepción es que la edición de las *Lecciones* se realizó formalmente hasta 1924, con ocasión del bicentenario del nacimiento de Kant, y estuvo a cargo del filósofo alemán Paul Menzer, quien publicó por primera vez una transcripción completa de las notas de los estudiantes de los cursos de ética de Kant (Heath & Schneewind, 1997, p. xiii[b]). Con todo, las *Lecciones de ética* no puede ser considerado un texto apócrifo. La investigación de Schmucker (1961; citado por Rodríguez Aramayo, 2012, p.17) ha demostrado el inmenso paralelo que existe entre el contenido de las *Lecciones* y el resto del pensamiento moral kantiano, certificando el carácter genuinamente kantiano de las *Lecciones de ética*[c].
La estructura interna de las *Lecciones* está asociada a dos libros de Baumgarten: *Initia philosophiae practicae primae* (1740) y *Ethica philosophica* (1751), ya que Kant tomaba estas dos obras como referencia para dictar sus clases de filosofía moral, pues en ellas se exponían suficientemente los problemas morales de la época. El mérito de Kant en el uso que hace de estos textos consistió en que, durante sus clases, él no repetía ni se ceñía al texto de Baumgarten, sino que iba comentando y entretejiendo sus propios puntos de vista éticos. Kant fue puliendo y labrando su filosofía moral durante la llamada "época del silencio" (1770-1780). A juicio de Rodríguez Aramayo, las *Lecciones* son el "laboratorio donde se gestó el formalismo kantiano" (2012, p.10) que será desarrollado más tarde en la *Fundamentación de la metafísica de las costumbres*, en la *Metafísica de las Costumbres* —especialmente en la segunda parte: *Principios metafísicos de la doctrina de la virtud*— y también en la *Crítica de la razón práctica.* Por esta razón, el valor filosófico de las *Lecciones* radica en que complementa y amplía muchas de las discusiones e ideas expuestas en las distintas obras de filosofía práctica de Kant, por ejemplo, se aprecia el rechazo al eudaimonismo, la idea del bien sumo (*summum bonun* [*höchsten Gut*]) y del bien supremo (*supremun bonum* [*oberstes Gut*]), la dignidad de ser feliz e incluso hay una incipiente versión del imperativo categórico.
[a] Las *Lecciones* están compuestas por las notas de clase de varias fuentes: Herder (1764-1765), Powalski (1777), Kutzner (1781), Th. Brauer (1780-1781), Collins (1784-1785), Mrogovius (1782) y Vigilantius (1793-1794). Existe una gran semejanza entre las distintas fuentes, lo cual ha llevado a pensar que Kant probablemente no alteró ni modificó radicalmente su material de clase durante el periodo en que fueron tomadas estas notas (1777-1791) (Heath & Schneewind, 1997, p. xv).
[b] La inclusión de las *Lecciones* en la edición de la Academia se realizó hasta 1974 —finalizó cinco años más tarde en 1979– y estuvo a cargo de Gerhard Lehmann. La traducción al inglés se realizó en 1930 por Louis Infeld. La versión al español se realizó hasta 1988 —con ocasión del bicentenario de la *Crítica de la razón práctica* (1788)— y estuvo a cargo de Roberto Rodríguez Aramayo y Concha Roldán Panadero.
[c] Es de resaltar que justamente las notas de Collins (1784-85) corresponden a la época en que Kant se encuentra redactando la *Fundamentación* (1785), razón por la cual es tan interesante comparar y en algunos

El texto de *Lecciones de ética* que presentaré a continuación data de 1785, y es anterior a "Sobre un supuesto derecho a mentir" (1797) y a la *Metafísica de las costumbres* (1797). En las *Lecciones*, Kant se esfuerza por explicar las formas en las que la moralidad nos ayuda a mejorar nuestro propio carácter, así como nuestras relaciones con los otros. Siendo fiel al deontologismo, Kant caracteriza estas relaciones en términos de *deberes*: deberes para con uno mismo, deberes para con otros hombres, y deberes para con los animales y los seres inanimados. Respecto de los deberes que tenemos hacia los otros seres humanos, Kant se detiene a hablar sobre el *deber ético de la veracidad* (*LE* 444-455). Lo llamativo es que allí Kant suscribe una posición en la que abre la posibilidad para mentir en caso de necesidad.

1.1.1. La mentira en caso de necesidad

En las *Lecciones*, como sucederá en la *Metafísica de las costumbres* y en "Sobre el supuesto derecho a mentir", Kant defiende la importancia de la veracidad en las declaraciones: "Manifestar las intenciones constituye la base principal de cualquier tipo de asociación humana y por eso es importante que cada uno sea sincero al comunicar sus pensamientos" (*LE* 444). A partir de esta idea se entiende también que la sinceridad tenga una función social indispensable: "Sin la sinceridad el trato social perdería su valor. No cabe instaurar sociedad humana alguna" (*ibid.*). "El mentiroso disuelve la congregación social ya que la mentira incapacita al hombre para sacar nada bueno del diálogo con su prójimo" (*LE* 445).

Sin embargo, la tesis más llamativa en las *Lecciones* es que Kant justifica la mentira. Según se afirma allí, si fuéramos sinceros siempre, entonces seríamos muy vulnerables a la maldad: "Si quisiéramos permanecer fieles en todos los casos a la exactitud de la verdad, quedaríamos con frecuencia a merced de la maldad de quienes pretendan hacer un mal uso de nuestra sinceridad" (*LE* 447). La sinceridad incondicional puede llegar a exponernos al mal o, en otras palabras, hacer el bien puede derivar en un mal.

---

casos complementar las dos obras.

La intuición general, tal y como reconoce Kant, consistiría en limitar la veracidad, o en otras palabras, consistiría en darle espacio a la posibilidad de mentir: "Tal y como son los hombres actualmente, es cierto que se corre a menudo peligro observando escrupulosamente la verdad y por ello se ha forjado el concepto de *mentira en caso de necesidad*, el cual constituye un punto muy delicado para un filósofo moral" (*LE* 447; énfasis añadido). Sin embargo, estas situaciones no pueden ser arbitrarias ni pueden ser la regla que reemplace al deber de la sinceridad: "Si se pudiera justificar el robo, el asesinato y la mentira por mor de la necesidad, el caso de necesidad vendría a sustituir a toda la moralidad y quedaría a juicio de cada cual estipular qué haya de considerarse como un caso de necesidad […] se tornarían inseguras las reglas morales" (*ibid.*).

La propuesta de Kant consiste, entonces, en darle sentido al concepto de mentira por necesidad sin que ello implique renunciar al deber de la sinceridad; en otras palabras, Kant justifica la mentira por necesidad limitando su aplicación para que la necesidad no "sustituya a toda la moralidad". ¿Cuándo entonces está justificado mentir por necesidad? Sobre este punto, Kant introduce un ejemplo: "Supongamos que al preguntarme alguien: '¿llevas dinero encima?', de no responder infiriera de mi actitud que lo llevo, si contestase afirmativamente, me lo quitará y, respondiendo que no, mintiese; ¿qué debo hacer en este caso?" (*LE* 448). La respuesta es sorprendente, pues a partir de este caso, Kant postula la posibilidad de la *mentira como defensa* (*ibid.*), y explicita muy claramente la situación en la que puede hacerse uso de ella:

> El único caso en que está justificado mentir por necesidad se produce cuando me veo coaccionado a declarar y estoy asimismo convencido de que mi interlocutor quiere hacer un uso impropio de mi declaración. (LE 448)

Para enteder esta afirmación habría que hacer una distinción importante: Kant distingue entre *falsiloquium* y *mendacium* (*LE* 446). Un *falsiloquium* se presenta en los siguientes casos: (i) cuando anuncio que voy a revelar mi intención y en lugar de eso hago una falsa declaración; (ii) cuando el otro no pueda presumir que estoy expresando mis intenciones (puedo engañar sin decir nada en absoluto); (iii) también incurro en *falsiloquium* cuando "albergo el propósito de ocultar mis intenciones a otros que pueda presumirlo así al proyectar él mismo hacer un mal uso de mi verdad, de este modo, si un enemigo me atenaza por el cuello y me pregunta dónde tengo el dinero, en tal caso puedo encubrir mis pensamientos, ya que él pretende hacer un mal uso de la verdad" (*LE*

447). En cambio, una *mendacium* es un tipo específico de falsedad (*falsiloquium*): "no toda falsedad es una mentira [*mendacium*]; lo es solo si uno ha declarado expresamente que quiere dar a entender al otro su propósito" (*LE* 448). El rasgo distintivo de la *mendacium* es que es un *falsiloquium* que perjudica a la humanidad en general,[4] porque al mentir rompemos el pacto de sinceridad con el otro y con ello "conculcamos el derecho de toda la humanidad" (*LE* 448).

Lo interesante de esta distinción es que incurrir en un *falsiloquium* no implica incurrir necesariamente en *mendacium*. Para el caso específico del ladrón que busca mi dinero, no es *mendacium* encubrir la verdad: "Esto no llega a ser *mendacium*, pues el otro sabe que disimularé mis pensamientos y que tampoco tiene derecho alguno a exigirme la verdad" (*LE* 447). Como se ve, existe una diferencia entre *encubrir* la verdad y *mentir*. El encubrimiento es un caso de *falsiloquium*, no de *mendacium*. Así que, en sentido estricto, no revelar el paradero de mi dinero no cuenta como una mentira realmente. Si atendemos a las características específicas de la situación, encontramos que el "encubrimiento de la verdad" tiene dos elementos relevantes. La primera es que el ladrón sabe o supone que no seré sincero y, segundo, el ladrón no "tiene derecho a exigirme la verdad". Estas dos características son muy importantes porque modifican la naturaleza del fenómeno: si el ladrón sabe que no soy sincero y si *además* no tiene derecho a exigirme sinceridad, entonces ya no cabe hablar de mentira. La mentira (*mendacium*), estrictamente hablando, requiere que la persona engañada en realidad confíe en lo que dice el mentiroso.

Se puede decir que cuando Kant justifica la mentira por necesidad en realidad está justificando un *falsiloquium* y no un *mendacium*. En otras palabras, la mentira por necesidad o mentira defensiva no es realmente una mentira, sino un encubrimiento de la verdad. Así que, en estricto sentido, Kant no está aceptando la posibilidad de la mentira, más bien, la conclusión correcta es que está justificado encubrir la verdad (*falsiloquium*) cuando se trata de un caso de necesidad.

Es más, Kant se reafirma en el deontologismo y sugiere que el agente *debe* mantenerse fiel al deber de la sinceridad. En las *Lecciones,* Kant refuerza esta tesis: "la mentira constituye algo indigno en sí

[4] "Jurídicamente la *mendacium* es un *falsiloquium in praejudicium alterius* —en este orden de cosas no podía ser de otra manera—, pero considerada moralmente se trata de un *falsiloquium in praejudicium humanitatis"* (*LE* 448).

misma, independientemente de que se tengan buenas o malas intenciones" (*LE* 448). La mentira es un incumplimiento de un deber hacia los demás y, como tal, es moralmente incorrecta. Incluso al final del ensayo de *Lecciones*, y en un tono muy optimista, Kant afirma que quien se atiene a la verdad no "dejará que tenga lugar ningún *casus necessitatis*", o mejor, no permitirá que haya *falsiloquium* justificados por necesidad:

> La mentira nunca reporta nada bueno. Un mentiroso es un cobarde que comienza a mentir porque no es capaz de conseguir algo ni tiene otra manera de salir de apuros. Por el contrario, una persona valiente amará la verdad y no dejará que tenga lugar ningún casus necessitatis. (LE 448)

1.2 "Sobre el supuesto derecho a mentir por amor a la humanidad"

En contraste con lo dicho en *Lecciones*, Kant suscribe posteriormente una posición mucho más radical con respecto a la mentira. El breve ensayo "Sobre el supuesto derecho a mentir por amor a la humanidad" (*SDMF*) (1797) es una respuesta a un texto del filósofo francés Benjamin Constant. Según Constant:

> El principio moral, por ejemplo, de que decir la verdad es un deber, si se tomase de manera absoluta y aislada, haría imposible toda sociedad. Tenemos la prueba de ello en las consecuencias muy directas que de ese principio ha sacado un filósofo alemán [se refiere a Kant, LG], que llega hasta a pretender que, ante asesinos que os preguntasen si vuestro amigo a quien persiguen se ha refugiado en vuestra casa, la mentira sería un delito. (SDMF 425)

¿Qué debemos hacer si un asesino que persigue a nuestro amigo nos pregunta por su paradero? ¿Estamos obligados a cumplir con nuestro deber de ser veraces y, por lo tanto, informarle a este hombre que nuestro amigo se encuentra en nuestra casa? ¿O dadas las posibles consecuencias violentas y negativas, estamos justificados a mentir?

Para ponerlo de modo general, lo cierto es que en "Supuesto derecho a mentir", Kant sostiene

taxativamente que la mentira es inadmisible en cualquier situación, incluso si la veracidad trae consecuencias negativas para nuestro amigo refugiado. El deber moral está por encima de cualquier inclinación o interés personal. Constant, en cambio, considera excesiva y equivocada la aplicación estricta de la veracidad, por eso aboga por un derecho a mentir en este tipo de circunstancias.

1.2.1 El deber de ser veraz ante el asesino en la puerta

"Sobre el supuesto derecho a mentir por amor a la humanidad[5]" se ha interpretado como una prueba de la insistencia kantiana en aceptar y cumplir las obligaciones morales de una manera rígida y extrema. Usualmente, este texto se ofrece como prueba de que la teoría moral kantiana conduce a resultados inaceptables moralmente. La razón de esta posición se explica porque allí Kant afirma categóricamente que "la veracidad en las declaraciones que no se pueden evitar es un deber formal del hombre con relación a cualquier otro, por mayor que sea el perjuicio que se deduce de esto para él o para otra persona" (*SDMF* 426). El problema de esta afirmación es que compromete a Kant con un fuerte rigorismo[6]: debemos acatar el deber de la veracidad sin importar

---

[5] La cronología del ensayo "Sobre un supuesto derecho a mentir por amor a la humanidad" es un poco confusa. Se sabe que fue escrito en 1797. Lo curioso es que en ningún otro pasaje de la obra de Kant hay una referencia al famoso ejemplo del asesino en la puerta. "Sobre un presunto derecho a mentir" fue publicado en septiembre de 1797. Por su parte, el texto de Constant fue escrito en mayo de 1796, pero K.Fr.Cramer lo tradujo y editó al alemán hasta 1797. Lo más probable es que el contacto entre Constant y Kant se debiera a Cramer, quien tradujo el ensayo de Constant al alemán y además fue el responsable de la publicación de la réplica de Kant. Al parecer, Constant se basó en un comentario de Cramer sobre la posición kantiana. También es posible que el ejemplo del asesino en la puerta criticado por Constant se haya debido a la influencia de J.D. Michaelis, un profesor de filosofía de Göttingen, quien al parecer sí sostenía la posición a la que Constant hace referencia. Aún así, Cramer sostiene que es Kant el "filósofo alemán" al que se refiere Constant. Por todo esto, es extraño que el propio Kant haya aceptado tener en cuenta una postura o un ejemplo como el del asesino en la puerta que en sentido estricto él nunca propone.

[6] El término "rigorismo" ha sido usado de dos maneras en la teoría moral sobre Kant (Shapiro, 2006, p. 33). En un sentido, se refiere al rigorismo con respecto a los deberes morales, es decir, a la exigencia de que los deberes morales perfectos son obligatorios y deben cumplirse incondicionalmente. En otro sentido, se refiere al rigorismo con respecto al carácter moral: "la idea de que el carácter de una persona debe ser considerado enteramente bueno o enteramente malo" (*R* 71/6:22). Para los propósitos de esta investigación, tendré en cuenta solo el primer sentido del concepto.

Una gran cantidad de autores han analizado el rigorismo con respecto a los deberes morales. En la siguiente

las circunstancias ni las consecuencias. Este resultado tan radical deja la sensación de que, en palabras de Christine Korsgaard, la teoría kantiana no permite responder ante el mal (2000b, p. 133) pues el deber se impone incluso si esto significa que las consecuencias son desastrosas para nosotros mismos o para otras personas.

Como decía anteriormente, Benjamin Constant encuentra la posición de Kant inaceptable. Su crítica está dirigida en contra de la *incondicionalidad* del deber. El cumplimiento irrestricto del deber "tornaría imposible cualquier sociedad", es decir, las consecuencias del rigorismo kantiano conllevarían a la destrucción de la sociedad. Por esta razón, Constant propone una versión *condicionada* del deber:

> El concepto de deber es inseparable del concepto de derecho. Un deber es aquello que corresponde en un ser a los derechos del otro. Donde no hay ningún derecho, no hay ningún deber. Por consiguiente, decir la verdad es un deber, pero solamente con aquellos que tienen derecho a la verdad. Ningún hombre, por tanto, tiene derecho a la verdad que perjudica a otros. (SDMF 425)

En este sentido, el asesino no tiene derecho a la verdad; él pierde este derecho cuando tiene como fin "perjudicar a otros"; y dado que para Constant el deber depende del derecho, nosotros no tenemos el deber de decir la verdad a quien no tiene el derecho a recibirla. Esto significa que, para Constant, la verdad no es ningún deber incondicionado; por el contrario, decir la verdad está condicionado a que nuestro interlocutor lo merezca. La fuerza del argumento de Constant está

---

lista de artículos puede verse un desarrollo cronológico y conceptual de este tema: Marcus Singer (1954) "The Categorical Imperative"; Paton (1954) "An Allegen Right to Lie: A Problem in Kantian Ethics"; Dietrichson (1964) "When Is a Maxim Fully Universalizable?"; Sally Sedgwick (1991) "On Lying and the Role of Content in Kant's Ethics"; Barbara Herman "Moral Deliberation and the Derivation of Duties" (1993a) y "Obligation and Performance" (1993b), ambos en *The Practice of Moral Judgment* (1993c)*;* Thomas E. Hill (1992) "Making Exceptions without Abandoning the Principle: Or How a Kantian Should Think about Terrorism"; Christine Korsgaard (1996) "Two Arguments against Lying" y "The Right to Lie: Kant on Dealing with Evil," ambos en: *Creating the Kingdom of Ends;* Allen Wood (2008b) "Lies". *Kantian Ethics.*
El rigorismo con respecto al carácter moral ha sido menos trabajado que el rigorismo con respecto a las normas morales. Una posible razón se debe a que el rigorismo del carácter se desarrolla en la *Religión dentro de los límites de la razón*, una obra que ha recibido menos atención que la *Fundamentación de la metafísica de las costumbres* o la *Crítica de la razón práctica*. Sobre el rigorismo del carácter, ver: Engstrom (1988) "Conditioned Autonomy"; Allison (1990) *Kant's Theory of Freedom*.

basada en una concepción *condicionada* de la veracidad: si bien tenemos el deber de ser veraces, este deber está condicionado a que el interlocutor *merezca* que lo hagamos o a que tenga el *derecho* ello.

La pregunta que tendría que responder una concepción condicionada de la verdad es bajo qué criterios se determina si un interlocutor tiene o no derecho a la verdad. A la luz del texto, se puede decir que el asesino ha perdido todo derecho a la verdad porque "nadie tiene derecho a la verdad que perjudica a otros". En mi interpretación, esto quiere decir que el asesino no tiene derecho a la verdad porque quiere hacer un *uso inapropiado* de ella. Para reforzar esta interpretación, no podemos olvidar, por ejemplo, que Constant mismo afirma que el "cumplimiento irrestricto del deber tornaría imposible cualquier sociedad", una frase que leo fácilmente en clave consecuencialista. En otras palabras, el derecho a la verdad también está *condicionado* a consideraciones *consecuencialistas*, pues el asesino pierde su derecho a la luz de las consecuencias negativas de sus actos que posiblemente cometerá con el conocimiento de la verdad.

Dicho esto, la concepción condicionada de la verdad depende de que el interlocutor tenga el derecho a la verdad, y a su vez, merecer este derecho depende de la evaluación de las posibles consecuencias de los actos realizados con conocimiento de la verdad. En breve, la postura de Constant, por un lado, prioriza el derecho por encima del deber (el derecho justifica el deber) y, por otro, introduce criterios consecuencialistas para evaluar la posesión de este derecho. Este análisis tal vez no sea suficiente para etiquetar a Constant como un consecuencialista, pues para eso se requeriría un esfuerzo adicional que no interesa por ahora. Basta con mostrar cómo se articula su argumento: el deber de la veracidad está condicionado a la posesión del derecho por parte del interlocutor y también está condicionado a la evaluación de las posibles consecuencias.

Kant, en contra de la postura de Constant, sostiene que el valor moral de la veracidad nada tiene que ver con la condición de los interlocutores ni con las posibles consecuencias de sus acciones. Para Kant, el deber de la veracidad es exigible independientemente de que nuestro interlocutor tenga derecho a la verdad, y también es exigible independientemente de las posibles consecuencias que traiga la verdad:

> Se define, por tanto, a la mentira como una declaración intencionalmente no verdadera hecha a otro hombre, y no hay necesidad de agregar que deba perjudicar a otra persona,

> como exigen los juristas en la definición que de ella presentan (mentir es falsificar en perjuicio de otro). Pues ella perjudica siempre a otra persona, incluso cuando no a otro hombre determinado y sí a la humanidad en general, al inutilizar la fuente del derecho. (SDMF 426)

Para Kant, la incorrección de la mentira nada tiene que ver con que se perjudique a otra persona en particular. Mentir es incorrecto, incluso si nadie en concreto resulta perjudicado. La incorrección de la mentira radica en otra parte. Para seguir el argumento de Kant, la mentira es incorrecta porque constituye la violación de un deber:

> [Cuando miento] sí cometo injusticia […] con la parte más esencial del deber en general: esto es, hago, en lo que de mí depende, que las declaraciones en general no encuentren creencia alguna y también con ello que caduquen y pierdan vigor todos los derechos que están fundados en contratos; lo cual es una injusticia cometida con la humanidad en general. (SDMF 426; énfasis en el original)

Este pasaje es difícil de interpretar, especialmente porque Kant introduce el concepto de 'deber en general'. Según mi interpretación, el deber que se viola es el de respetar la idea misma de comunicación humana. Este es un deber general porque la comunicación es común a todo tipo de relación con otros. "Manifestar las intenciones constituye la base principal de cualquier tipo de sociedad y por eso es tan importante que cada cada cual sea sincero al comunicar sus pensamientos" (*LE* 444). Irrespetar la base misma de la comunicación humana implica obstaculizar e incluso impide la relación con otros porque "la mentira incapacita al hombre para sacar nada bueno del diálogo con su prójimo" (*ibid.*). El problema con la mentira es que a través de ella se estaría contrariando el fin de la comunicación, que es que los hablantes manifiesten adecuadamente los pensamientos entre sí. Si la mentira fuera la norma ya no habría razón para creer en las declaraciones en general. Y de paso, también se deslegitima todo tipo de contrato porque no es posible pactar si no se confía en que el otro está siendo veraz. "Solo se puede juzgar lo que el otro piensa a partir de sus declaraciones, las cuales han de ser fiel expresión de sus pensamientos, pues *sin ese presupuesto no cab*e *instaurar sociedad humana alguna*" (*LE* 444; énfasis añadido).

Si esta interpretación es correcta, entonces ya es más claro por qué la mentira es incorrecta a la luz de una intuición deontológica: mentir constituye una violación al deber en general porque perjudica a la humanidad en su totalidad, pues con la mentira socavo los cimientos que sirven de base a cualquier tipo de asociación humana. A partir de esto, Kant concluye con su más fuerte tesis:

> Es por tanto un sagrado mandato de la razón, que ordena incondicionalmente y no admite limitación, por cualquier especie de conveniencia, lo siguiente: ser veraz en todas las declaraciones. (SDMF 427)

Si lo analizamos detenidamente, la discusión Constant-Kant enfrenta dos intuiciones morales: la intuición condicionada de la veracidad y la intuición deontológica. La primera intuición nos indica que mentirle al asesino está justificado porque el asesino no podría exigir el deber de la veracidad cuando no tiene ningún derecho ella. Por otro lado, la intuición deontológica nos indica que la mentira es una acción inmoral por sí misma porque supone una violación al deber en general. Traducido en acciones concretas, esto significa elegir entre ser veraces en nombre del deber, o mentir dado que el asesino no merece la verdad.

1.2.2 Contexto histórico-político de "Sobre un supuesto derecho a mentir"

La posición rigorista que expresa Kant en "Sobre un supuesto derecho a mentir" ha sido blanco de múltiples críticas. Para muchos, este resultado es la prueba contundente de que la teoría kantiana es insostenible. Desde luego, como veremos más adelante, tampoco han sido pocos los intentos por tratar de salvar o matizar la posición kantiana. Por ahora, me interesa analizar una de las interpretaciones favorables a Kant que propone abordar "Sobre el supuesto derecho a mentir" desde una perspectiva política e histórica. Según Benton (1982), este breve ensayo es una respuesta a un panfleto político de Benjamin Constant titulado "De las reacciones políticas" [*Des réactions politiques*] publicado en alemán en 1797.

El contexto de la discusión Constant-Kant, dice Benton, se generó a propósito de que el Directorio Francés se proclamara como el heredero de los verdaderos valores de la Revolución Francesa, al

ganar el poder en Francia en contra de los jacobinos en 1795 (Benton, 1982, p. 137). El Directorio fue creado como una reacción contra las tendencias más radicales de la Revolución; específicamente, se oponía al radicalismo de los jacobinos. Dentro de este contexto, el panfleto de Constant fue escrito para justificar las políticas del Directorio: "profesar un firme compromiso con los principios, pero al mismo tiempo su objetivo primario es prevenir el exceso al establecer límites estrictos en la aplicación de los principios" (Benton, 1982, p.137). El panfleto de Constant puede leerse entonces como un texto político anti-jacobino, pues es bien conocido el hecho de que el terror impuesto por los jacobinos fue el resultado directo de "aplicar principios abstractos a situaciones de la vida real" (*ibid.* p. 138).

Por su parte, Kant siempre estuvo comprometido públicamente con los principios de la Revolución Francesa, incluso continuó defendiéndola durante la época del Terror (Benton, 1982, p. 138). Por esta razón, Kant se ganó la reputación de jacobino. A la luz de esta idea, cuando Constant se refiere a Kant ("el filósofo alemán") y retoma el ejemplo de la mentira, lo hace dentro de un contexto, podríamos decir, panfletario, para mostrar que "si los principios son tomados muy estrictamente, pueden conducir a excesos y a la destrucción de la sociedad […]. El significado real del ejemplo del asesino es su papel en un argumento más general para justificar la flexibilidad de los principios en nombre de la preservación de la sociedad" (*ibid.*). Desde esta perspectiva, Kant parece haber entendido el panfleto de Constant como un texto con pretensiones políticas que, en su opinión, atacaba los principios de la Revolución Francesa. En otras palabras, Kant encontró que el objetivo de Constant "por desacreditar el rigor de los principios" (*ibid.* p. 137) iba dirigido a apoyar políticamente al Directorio.

Para Kant, la posición de Constant resulta políticamente muy problemática. De allí que escriba su réplica pensando en un contexto político, más específicamente, pensando en la Doctrina del Derecho y no tanto en un problema ético. Por esta razón, la tesis de Benton es que "Sobre el supuesto derecho a mentir" es un ensayo que no aborda un problema ético, sino un problema de derecho [*Recht*][7] (político). Y en esta medida, según Benton, las conclusiones de Kant deben leerse

[7] Benton es muy insistente con el sentido del término *Recht*. "He elegido no traducir el término *Recht* de Kant, ya que creo que ni 'justicia', ni 'ley', ni 'derecho' son una traducción adecuada para el significado que Kant pretende aquí" (1982, p. 136n7). Más adelante aclara: "Me gustaría sugerir que el problema de *Recht* que se plantea aquí no es simplemente un problema de justicia (en el sentido de derecho positivo o en

desde la teoría del derecho y no como una prescripción ética.

Ahora bien, como le mencionaba anteriormente, el argumento de Constant se basa en una lectura condicionada de los deberes: "donde no hay derechos no hay deberes". Esta es otra manera de decir que *los derechos preceden o condicionan a los deberes*. Probablemente Constant no era consciente de la profundidad de esta afirmación (Benton, 1982, p. 140), y tal vez solo la introdujo para sustentar su argumento contra el deber incondicionado de la veracidad. Lo cierto es que, para Kant, esta afirmación sí tiene implicaciones profundas para la teoría política con las que no está de acuerdo.

La principal implicación de afirmar que los derechos preceden a los deberes es que avala una perspectiva hobbesiana de la política que Kant rechaza explícitamente[8]. Para Hobbes, la subordinación de los deberes a los derechos es una característica propia de una sociedad bien organizada y además es una descripción adecuada de la psicología práctica de los seres humanos en el Estado de Naturaleza. En efecto, Hobbes propone una antropología en la que los seres humanos solo son motivados a actuar si hacerlo comporta un beneficio o un placer, o si se evita un dolor o un mal.[9] De esta manera, quienes suscriben el contrato adquieren obligaciones y deberes porque desean preservar o asegurar ciertos derechos (p.e. a la vida o la propiedad) que no están garantizados en el Estado de Naturaleza. En este sentido, los deberes se adquieren y se validan con vistas a preservar privilegios o derechos, es decir, los derechos condicionan y, por tanto, preceden a los deberes. Ninguna obligación puede existir para Hobbes excepto dentro de un Estado constituido, y es por ello que los deberes están subordinados a los derechos, los derechos preceden a los deberes: "los deberes surgen solo dentro del contrato social, y el único fundamento para la necesidad del contrato en sí mismo es evitar la coacción física y la inseguridad material. Por la misma razón, no puede haber derechos o deberes que sean más básicos que aquellos proclamados por el estado […] excepto el derecho a preservar la propia vida y la propiedad" (Benton, 1982, p.

---

el sentido de los principios en los que debería basarse el derecho positivo) sino más bien de *Recht* en cuanto se aplica a la política (1982, p. 136s).

[8] Kant desarrolla una crítica puntual a Hobbes en la Parte II de *Teoría y Práctica: "De la relación entre teoría y práctica en torno al Derecho político (Contra Hobbes)"*.

[9] En la Primera parte del *Leviatán* (1651) Hobbes se esfuerza por presentar una teoría mecánica de las pasiones. La antropología hobbesiana concibe al ser humano como un ser dotado de movimiento dentro de un mundo de causas. De manera que la motivación que tenga un agente hobbesiano también se explica en términos causales: tiende a aquello que le cause agrado o placer, y rehuiría aquello desagradable.

140).

La teoría política de Kant es completamente opuesta a la perspectiva hobbesiana[10]. Para empezar, la antropología kantiana supone agentes que son capaces de actuar independientemente de la relación entre causas y determinaciones que constituyen al mundo sensible (§3.4); en otras palabras, supone agentes que pueden actuar no por inclinación ni mecánicamente, sino libremente (§3.2). Aquí ya se marca una distancia en relación con la antropología hobbesiana, pues los motivos de la acción —para Hobbes—, al menos en un estado de naturaleza, solo pueden ser el placer y el dolor (inclinaciones); para Kant, en cambio, las máximas racionales tienen y deben tener capacidad motivacional, y actuar libremente —al menos en sentido negativo (*F* 4:446)— es actuar sin influjo de inclinaciones ni pasiones[11].

Para Kant, la libertad —en sentido positivo (*F* 4:446)— se expresa en las acciones motivadas por deber moral, es decir, por acciones determinadas por la razón práctica (leyes morales). Y adicionalmente, en la *Doctrina del derecho*, Kant establece que el principio fundamental [*Grund*] del derecho [*Recht*] deriva de la ley moral (*MC* 230-31[12]). Por tanto, el razonamiento concluye que el deber, en tanto que ley moral, precede lógicamente a todos los derechos y es su condición de posibilidad. A diferencia de Constant y de Hobbes, Kant considera que los deberes son anteriores a los derechos.

Con esto dicho, el ejemplo del asesino en la puerta adquiere un nuevo significado. Desde una lectura histórica-política del ensayo, este ejemplo concreto no es lo relevante, más bien, lo que está de fondo es el importante papel de la veracidad en la constitución de cualquier sociedad humana.

---

[10] Ver: *Teoría y práctica*, Parte II; *La Paz perpetua (PP)*, Apéndice I (G375).

[11] Kant analiza el influjo de las pasiones, emociones e inclinaciones sobre la voluntad humana en el Tercer libro de la *Antropología en sentido pragmático* (§73ss). La relación antagónica entre las inclinaciones y la libertad se analiza en la *Fundamentación* (*F* 4:446).

[12] En Doctrina del Derecho —primera parte de la *Metafísica de las Costumbres*—, Kant dice: "Una acción es conforme a derecho [*Recht*] cuando permite, o cuya máxima permite a la libertad del arbitrio de cada uno coexistir con la libertad de todos según una ley universal" (*MC* 230). De modo general, de aquí se desprende que las acciones permisibles lo son en virtud de que no contradicen la libertad de otros seres racionales, siempre que esa libertad se entienda como la capacidad del agente para autodeterminarse racionalmente. Como se ve, el derecho (específicamente, las acciones conforme al derecho) supone toda la teoría de la libertad kantiana, y en ese sentido, se explica la dependencia entre derecho y moralidad.

Bajo esta lectura, la respuesta de Kant a lo dicho por Constant tiene el siguiente sentido:

> Visto políticamente en términos de la teoría del contrato social, el deber de la veracidad es uno de los deberes más fundamentales de todos los deberes: debe preceder lógicamente incluso al contrato social en sí mismo ya que la obligación de ser veraz es una presuposición de cualquier contrato justificable. (Benton, 1982, p. 141)

Esta es la razón por la que el deber a la veracidad es anterior y más fundamental que el supuesto derecho a mentir al asesino, y también es más fundamental que cualquier tipo de derecho que tenga el asesino a exigirnos una respuesta veraz.

Al poner la discusión en términos de teoría política, no hacemos referencia a la veracidad en un contexto de comunicación cotidiana, sino al hecho de que ella es una condición para el establecimiento de una sociedad humana; de ahí que "la mentira socava las bases mismas de la sociedad" (*SDMF* 427). A la luz de este contexto político, la respuesta de Kant resulta menos escandalosa. Cuando Kant asegura que debemos ser veraces con el asesino, está pensando en la dimensión política de la veracidad y no en la dimensión ética de enfrentar cara a cara al asesino.

Esta respuesta es interesante porque ubica la posición de Kant en un contexto donde ya no resulta tan problemática. Sin embargo, al recontextualizar la posición de Kant en el ámbito de la teoría política y del Derecho, queda todavía un interrogante ético, algo que denomino un *vacío ético*: puede que sea cierto que la verdad es un valor fundamental para establecer una sociedad humana, pero no es obvio que esta exigencia incondicional de veracidad deba extenderse al campo ético. ¿Qué debemos hacer *éticamente* frente al asesino en la puerta? ¿Acaso la fuerza normativa de la veracidad como principio político es la misma para la veracidad como deber hacia los otros? ¿Por qué Kant se serviría de un ejemplo éticamente relevante para dar cuenta de un principio político?

1.3 *Metafísica de las costumbres*

La aproximación política-histórica a "Sobre el supuesto derecho a mentir" aparentemente neutraliza la acusación de rigorismo que pesa en contra de Kant. Con todo, quien acoge esta

interpretación deja de lado que Kant también trató el tema de la mentira y la veracidad desde una perspectiva exclusivamente ética, de manera que el interrogante por cómo responder éticamente al asesino tiene, en efecto, una respuesta dentro de la filosofía práctica kantiana. Esta respuesta, al menos como la interpreto, puede extraerse de la *Metafísica de las costumbres* (1979). Lo interesante de este último texto es que llega a la misma conclusión de "Sobre el supuesto derecho a mentir": la mentira es incorrecta porque constituye una violación de un deber. Por supuesto, con una batería argumentativa diferente. De modo que, incluso si resulta plausible la interpretación política-histórica que Benton hace de "Sobre el supuesto derecho a mentir", todavía queda en pie el rechazo, en sentido *ético*, de cualquier forma de mentira. Es posible concluir entonces que Kant es consistente con el rechazo a la mentira, ya sea en términos políticos o en términos éticos.

1.3.1 Deberes hacia uno mismo

En la *Doctrina de la virtud* —segunda parte de la *Metafísica de las costumbres*— Kant afirma que mentir comporta una violación de los deberes para con uno mismo (*MC* 429). A primera vista puede resultar extraño que existan deberes hacia uno mismo, pues se tiene la idea de que los deberes son acciones orientadas hacia otras personas. ¿Cómo entonces es posible tener deberes u obligaciones morales con uno mismo? Kant es consciente de esta dificultad, e incluso considera la posibilidad de que el concepto de deber hacia uno mismo sea contradictorio:

> Si el yo que obliga se toma en el mismo sentido que el yo obligado, el deber hacia sí mismo es un concepto contradictorio. Porque en el concepto de deber está contenido el de una coerción pasiva (yo soy obligado). Ahora bien, como es un deber hacia mí mismo, me represento como obligando, por tanto, ejerciendo una coerción activa (yo, el mismo sujeto, soy el que obliga); y la proposición que expresa un deber hacia sí mismo (yo debo obligarme a mí mismo) contendría una obligación de estar obligado (una obligación pasiva que, sin embargo, sería a la vez activa en el mismo sentido de la relación), por consiguiente, contendría una contradicción. (MC 417; énfasis en el original)

A pesar de esta supuesta contradicción, los deberes hacia sí mismo deben existir, ya que "si no hubiera tales deberes [hacia sí mismo], no habría ningún deber en general" (*MC* 417). Los deberes

hacia uno mismo son condición de posibilidad de la idea misma de deber:

> [P]orque yo no puedo reconocer que estoy obligado a otros más que en la medida en que me obligo a mí mismo: porque la ley, en virtud de la cual me considero obligado, procede en todos los casos de mi propia razón práctica, por la que soy coaccionado, siendo a la vez el que me coacciono a mí mismo. (*MC* 417-18)

¿Cómo entender entonces el concepto de deber hacia sí mismo? La solución kantiana a la posible contradicción que comporta la idea de un deber hacia uno mismo apela a la expresión *punto de vista [Standpunkt]* (*F* 4:450; 4:452; 4:455; 4:458). El ser humano, cuando tiene conciencia de un deber hacia sí mismo, debe considerarse desde dos puntos de vista distintos (ver §3.3): como ser sensible (*homo phaenomenon*) y como ser racional (*homo noumenon*). Como seres sensibles, los seres humanos pueden verse determinados a partir de causas, en la medida en que realizan acciones en el mundo sensible; pero desde el punto de vista de su personalidad, como dotados de libertad, los seres humanos se consideran como capaces de verse obligados, particularmente, como capaces de obligarse a sí mismos. Estamos obligados en virtud de la condición de humanidad que constituye nuestro ser personas. Es por respeto [*Achtung*] a su propia condición de humanidad que el ser humano adquiere conciencia de los deberes u obligaciones hacia sí mismo. "De modo que el hombre (considerado en el doble sentido) puede reconocer un deber hacia sí mismo, sin caer en contradicción consigo mismo (porque no se piensa el concepto de hombre en uno y el mismo sentido)" (*MC* 418).

1.3.2 La mentira como violación del deber hacia uno mismo

Esta aclaración sobre los deberes hacia uno mismo es relevante en el contexto de esta investigación porque Kant afirma que la mentira es un vicio que se opone al deber del hombre para consigo mismo: "La mayor violación del deber del hombre para consigo mismo, considerado únicamente como ser moral (la humanidad en su persona), es lo contrario de la veracidad: la mentira" (*MC* 429).

Para entender mejor esta afirmación habría que partir de la distinción kantiana entre personas y cosas. Una primera diferencia es que las cosas son medios y sirven como instrumento para algún

fin adicional; las personas, en cambio, no son medios, sino que son fines en sí mismas (*F* 4:434). Una segunda diferencia es que una cosa es reemplazable, su valor es relativo, conmensurable e intercambiable con otra cosa y, en ese sentido, las cosas tienen un precio. En cambio una persona no es reemplazable, tiene un valor intrínseco, es decir, posee dignidad (*F* 4:434). Una tercera diferencia es que las cosas, al ser medios, están sometidas a una cadena causal medios-fines; en cambio las personas son causas primeras, no son medios, sino que, en sentido kantiano, son libres (tanto en sentido negativo como positivo) (*F* 4:439). Kant además sostiene que todos los seres humanos, en tanto que son seres racionales, cumplen con la condición de ser personas. En este sentido, hay una suerte de equiparación entre la condición de humanidad y la condición de persona. De manera que tratar a alguien como persona supone respetar su condición y valorar la humanidad en ella.

Si para Kant la mentira es un vicio que se opone al deber del hombre hacia sí mismo, ello se debe a que el mentiroso ha atentado contra el valor de la humanidad en su propia persona. Al mentir ha deshonrado o irrespetado su condición de persona, él mismo ha atentando contra su condición.. Con la mentira nos convertimos en objeto de desprecio ante nuestros propios ojos: "atentamos contra la dignidad de la humanidad en nuestra propia persona" (*MC* 429). Y la razón por la que la mentira nos deshonra ante nuestros propios ojos es que el mentiroso está actuando con un "fin opuesto a la finalidad natural de su facultad de comunicar sus pensamientos, por tanto, es una renuncia a su personalidad" (*MC* 429). En otras palabras, uno de los rasgos propios de la condición humana es la racionalidad, y esta racionalidad se manifiesta en distintas facultades, una de ellas, por supuesto, es la "facultad de comunicar pensamientos" (que incluye pensar, tener creencias, hablar). Así que ir contra el fin natural de esta facultad supone atentar contra algo muy característico y muy propio de nuestra capacidad racional y, por tanto, supone un desprecio a nuestra humanidad y a nuestra personalidad.

La mentira entonces no solo comporta una violación del deber hacia uno mismo. Ella es también un falta moral contra las otras personas. Esta última tesis no se encuentra explícitamente desarrollada en la *Metafísica de las costumbres*, y aun así considero que puede fácilmente derivarse a partir de la diferencia entre cosas y personas. Y es que si se piensa bien, el mentiroso convierte a la víctima en una cosa para sus fines; en otras palabras, el mentiroso trata como medio a la víctima:

la razón de la víctima es usada como medio para otra cosa.

> El que hace una promesa falsa piensa: si le digo que le pagaré la próxima semana escogerá darme el dinero. Esa persona está operando entonces con tu razón como con una máquina: la persona que engaña trata de determinar qué palancas accionar para obtener de ti los resultados deseados. La coerción física trata a la persona como una herramienta; la mentira trata a la razón de alguien como una herramienta. (Korsgaard, 2011, p. 280; énfasis en el original)

La mentira es una disposición a través de la cual estamos usando la razón de la víctima al servicio de las intenciones del mentiroso, le estamos negando la posibilidad a la víctima de hacer ejercicio de su propia racionalidad. "Cualquier intento de controlar las acciones y las reacciones por cualquier medio que no sea una apelación a la razón trata a la persona como un mero medio, dado que con ello se intenta reducirla a una causa mediata" (Korsgaard, 2011, p. 282). Además, en la medida en que la razón de la víctima de la mentira está siendo usada como medio, la víctima está perdiendo control sobre su capacidad de decidir y también sobre su capacidad de actuar. La mentira afecta nuestras creencias, y por lo tanto, afecta también nuestras deliberaciones a la hora de decidir o actuar. "El engaño interfiere con el derecho de las personas a gobernar su propio comportamiento" (Cholbi, 2013, p. 442). Como vemos, a la luz de estas ideas, resulta claro el sentido en el que la mentira y el engaño suponen tratar al otro como un simple medio.

Ahora bien, si mentir instrumentaliza a la víctima, ello implica atentar también contra su dignidad[13]. Y atentar contra la dignidad humana constituye una violación del deber moral porque usar al otro solo como medio es una acción contraria a la Formulación de la Humanidad del Imperativo Categórico.[14] Y toda máxima de acción que vaya en contra del Imperativo Categórico

---

[13] Korsgaard lleva esta línea argumentativa más lejos y considera que la mentira no solo afecta la dignidad de la víctima, sino que representa una "violación directa de la autonomía" (Korsgaard, 2011, p. 280). La idea que está de fondo en esta interpretación es que al tratar a la víctima como cosa, se niega su libertad, y en esta medida, se niega también su autonomía.

[14] Más adelante haré una exposición acerca del Test del Imperativo Categórico y sus distintas formulaciones (Ver §2.1). Por ahora, basta decir que la Formulación de la Humanidad del Imperativo Categórico nos exige que veamos al otro "siempre al mismo tiempo como fin y nunca simplemente como medio" (*F* 4:428). Si únicamente usáramos a ese otro como medio, entonces nuestra máxima de acción sería inmoral, estaríamos negando su dignidad y tratándola como una cosa.

representa una violación del deber moral. De modo que la mentira, al ser contraria al imperativo moral, no puede más que constituir una violación al deber moral.

La mentira, a la luz de la *Metafísica de las costumbres*, es entonces un vicio que atenta contra el deber hacia uno mismo y también un acto que supone una violación al deber moral hacia los otros. Es importante aclarar que en el contexto kantiano la mentira no es un vicio por los efectos nocivos que pueda traerle al mentiroso o a la víctima. Por ejemplo, la mentira, cuando es descubierta, nos deja mal parados frente a los otros; y esta mala reputación traería una consecuencia indeseable para el mentiroso, pues sus palabras ya no serían tomadas en serio, ya nadie confiaría en lo que dice. A Kant, por lo visto, le tienen sin cuidado las consecuencias de la mentira; y en vez de ello, como he tratado de mostrar, sostiene que la mentira es incorrecta porque es una violación del deber moral. Otra forma de expresar esta misma idea es sosteniendo que la prohibición de mentir no se justifica por consideraciones prudenciales[15] o por el auto-interés, sino porque representa una violación a la dignidad (propia y ajena) y es, por tanto, una transgresión contra los deberes para con uno mismo y una transgresión contra la dignidad de las víctimas.

He propuesto este análisis con el propósito de responder el interrogante ético del asesino en la puerta. Y es que el mismo Kant expresa que, sin importar el fin que tengamos, incluso si se trata de un fin bondadoso o meritorio como salvar a nuestro amigo, lo cierto es que la mentira es intrínsecamente incorrecta. Al mentir, sin importar las circunstancias específicas en que nos encontremos, estamos tratando al otro instrumentalmente y le estamos arrebatando su condición de persona.

La mentira (en el sentido ético de la palabra), como falsedad deliberada, no precisa

[15] El concepto de 'prudencia' [*Klugheit*] es un término técnico en la filosofía práctica kantiana. La prudencia es "la habilidad para elegir los medios relativos al mayor bienestar propio […]. El término 'prudencia' admite una doble acepción según nos refiramos a la 'prudencia mundana' o a la 'prudencia privada'. La primera es la habilidad que posee un hombre para tener influjo sobre los demás en pro de sus propósitos. La segunda es la pericia para hacer converger todos esos propósitos en pro del propio provecho duradero" (*F* 4:416n). En este sentido, la prudencia está restringida a la consecución de la felicidad propia —no al cumplimiento de la moral. De ahí que toda máxima prudencial estará siempre regida bajo el imperativo hipotético asertórico que se refiere a la elección de los medios para la felicidad. Las máximas derivadas de este imperativo asertórico son llamadas, justamente, *consejos de la prudencia,* pues ellas conciernen al bienestar y la felicidad de quien los sigue (*F* 4:416).

> perjudicar a otros para que se le considere reprobable; porque en ese caso la mentira sería violación de los derechos de otros. Su causa puede ser la ligereza o la bondad, incluso puede perseguirse con ella un fin realmente bueno, pero el modo de perseguirlo es, por la mera forma, un delito del hombre contra su propia persona y una bajeza que tiene que hacerle despreciable a sus propios ojos. (MC 430)

A partir de este pasaje, creo, se puede decir que es un deber ser veraces ante el asesino en la puerta. No solo porque actuando contra este deber transgrediríamos los deberes hacia nosotros mismos, sino porque además estaríamos mintiendo amparados en las posibles consecuencias bondadosas. Dado que estas posibles consecuencias son irrelevantes moralmente y en cambio el respeto a la humanidad en nuestra propia persona sí es relevante,[16] entonces lo que debemos hacer es ser veraces con el asesino.

1.3.3 ¿Qué hacer frente al asesino en la puerta?

Hasta este punto he analizado la evolución de la posición kantiana con respecto a la mentira. En las *Lecciones de ética* se abre la posibilidad de mentir por necesidad, aunque en sentido estricto, se trata más bien la posibilidad de encubrir la verdad (*falsiloquium*) ante la circunstancia en que la verdad sea mal usada. El encubrimiento de la verdad, según lo dicho por Kant en las *Lecciones,* no constituye necesariamente una mentira. "Sobre un supuesto derecho a mentir" y la *Metafísica de las costumbres,* en cambio, muestran un rechazo categórico y rigorista en contra de la mentira. La diferencia estriba en el contexto en el que es analizado el fenómeno de la mentira. "Sobre un supuesto derecho a mentir", si hemos de creerle a Benton, responde a una problemática histórica y política, y tiene como trasfondo la teoría política kantiana. Lo que se analiza en este texto es la relación de prioridad que debe haber entre deberes y derechos en el contexto de cualquier sociedad humana. En todo caso, este texto, incluso si se lee en sentido político, nos deja con la impresión de que Kant es un rigorista. Además el texto abre un interrogante ético que no se puede dejar de lado una vez planteado: ¿qué debemos hacer éticamente frente al asesino en la puerta? ¿Cuál es calidad ética de la mentira? Para responder estas preguntas me he servido de la argumentación de la

[16] Shapiro (2006) incluso ha llegado a sostener que mentirle al asesino para evitar que cometa el crimen es violar su libertad, aunque sea una libertad encaminada al mal. Mentir al asesino para protegerlo es una suerte de paternalismo que atenta contra la libertad y la autonomía. Esta línea de argumentación, por ahora, me lleva más lejos de mis intereses. Más adelante volveré sobre este punto (§2.4).

Metafísica de las costumbres donde se lleva a cabo una aproximación ética a la mentira, y en donde se concluye que *cualquier* mentira es una violación de un deber hacia uno mismo y también una violación hacia los otros en la medida en que ella instrumentaliza a la persona engañada.

A primera vista parece haber un aire de contradicción entre las *Lecciones* y los otros dos textos. Sin embargo, esta contradicción es aparente. En realidad, como he insistido, lo que se justifica en las *Lecciones* es el *encubrimiento de la verdad* ante situaciones en las que la verdad será mal usada (léase, en situaciones que conlleven consecuencias perjudiciales), pero esto no significa justificar la mentira. Y en este sentido, creo, es posible sostener que la posición de Kant en *Lecciones* en realidad no representa una contradicción para la tesis que va a desarrollar en "Supuesto derecho a mentir" y en la *Metafísica de las costumbres,* claro, a condición de aceptar que encubrir la verdad no equivale a mentir.

Sin embargo, esta interpretación tiene una implicación importante, y es que la noción de encubrimiento de la verdad (mentira por necesidad) sería perfectamente aplicable al caso del asesino en la puerta. En el ejemplo presentado por Constant es claro que el asesino desea hacer un mal uso de la información sobre el paradero del amigo escondido. Concuerdo con Rodríguez Aramayo en que las condiciones que Kant estipula en el ejemplo del ladrón que busca el dinero son exactamente iguales a las condiciones del caso del asesino en la puerta: en las dos situaciones la persona está coaccionada a declarar sabiendo que el "interlocutor quiere hacer un uso impropio de la declaración" (Rodríguez Aramayo, 2012, p. 275n). Tanto el ladrón, como el asesino han de suponer que no seré sincero porque la sinceridad se trastoca en un contexto de coacción, y además ninguno de ellos tiene derecho a exigir la verdad porque, como bien lo dijo Constant, este derecho se pierde cuando se tiene la intención de hacer un mal uso de la verdad.[17] Si esto es así, el caso del asesino justificaría perfectamente el encubrimiento de la verdad y sería, por tanto, un caso de

[17] Al hacer una comparación detenida entre las *Lecciones* y "Supuesto derecho a mentir" encuentro bastante extraño que Kant, en el primer texto, sostenga que el ladrón no tiene derecho a exigir la verdad, pero en el segundo texto arremete en contra de Constant, justamente porque este afirma que el asesino no tiene derecho a exigir la verdad. Evidentemente aquí hay una contradicción: Kant acepta y niega la misma tesis. Tal vez la tesis sea válida en términos éticos, pero no en términos políticos; o tal vez Kant solo haya cambiado de opinión. En cualquier caso, sería interesante analizar esta inconsistencia. En esta investigación no me detendré en este punto ya que este primer capítulo tiene un carácter más expositivo que argumentativo.

falsiloquium. ¿Por qué entonces no estaríamos justificados a mentirle por necesidad al asesino en la puerta?

Aunque no haya una contradicción conceptual entre el rigorismo y el concepto de mentira por necesidad, sí que sigue abierto el interrogante ético: ¿cuál es el curso de acción éticamente relevante frente al asesino en la puerta? Escudados en las *Lecciones*, podemos encubrirle la verdad al asesino, pero atendiendo a la *Metafísica de las costumbres* y a "Supuesto derecho a mentir" debemos de ser veraces. Lo que aquí tenemos es una tensión entre una aproximación universalista y otra casuística. La *Metafísica de las costumbres* y "Supuesto derecho a mentir" son textos en los que se hace un análisis abstracto y conceptual sobre la mentira. *Lecciones de ética,* en cambio, exhibe un "descenso a los terrenos de la casuística" (Rodríguez Aramayo, 2012, p. 25) con respecto a la mentira. En este caso, cuando la mentira se piensa no en términos universales, sino en una situación más específica, dentro de circunstancias concretas, los análisis conceptuales y abstractos se quedan cortos. Aun cuando no hay contradicción conceptual, eso no resuelve la pregunta práctica desde la perspectiva kantiana: ¿qué hacer frente al asesino en la puerta?

Todo este primer capítulo ha sido un ejercicio expositivo. Me he concentrado solo en reconstruir los argumentos kantianos sobre la mentira en los diversos textos en los que este tema ha sido tratado por el autor. Este camino es requerido porque ofrece una fotografía conceptual que será relevante para caracterizar el rigorismo moral kantiano. Sin embargo, el análisis, como hemos visto, no ha sido del todo concluyente. En el siguiente capítulo, abordaré el caso del asesino en la puerta teniendo en cuenta otro enfoque: considero que apelando a la noción de deberes perfectos es posible responder al interrogante ético formulado a partir del caso del asesino en la puerta. Por otra parte, y además, un análisis de esta noción sirve bien a mi propósito de defender la posición según la cual la filosofía práctica de Kant es rigorista.

2. Deberes perfectos y rigorismo kantiano

2.1 Acción por deber y acción conforme al deber
- 2.1.1 Buena voluntad y deber
- 2.1.2 Acción por deber y acción conforme al deber
- 2.1.3 Sobredeterminación motivacional [*motivational overdetermination*]
- 2.1.4 Imperativo como deber-ser

2.2 Imperativo categórico
- 2.2.1 El procedimiento del imperativo categórico

2.3 Deberes perfectos y deberes imperfectos

2.4 La máxima de mentirle al asesino

2.5 La mentira: usar al otro como medio

2.6 Korsgaard: el derecho a mentir

La razón por la que me he concentrado en entender la posición de Kant sobre la mentira es porque es indesligable de la acusación de rigorismo que pesa en contra de su filosofía moral. Creo que esta acusación es correcta, o eso al menos es lo que intentaré defender en este capítulo. Más adelante desarrollaré el concepto de rigorismo (cap. 3), pero por lo pronto, este calificativo se referirá a la exigencia del cumplimiento incondicional del deber de la veracidad. El análisis que desarrollé en el capítulo anterior muestra que existe evidencia textual para sostener que Kant hace una interpretación rigorista del caso presentado, aunque también es verdad que él puede llegar a admitir que estamos justificados a mentir por necesidad. Lo que intenté exponer allí era un ejemplo de la tensión que se da entre la exigencia de incondicionalidad de los deberes morales y la aplicación concreta de estos deberes. Las *Lecciones de ética* de Kant, con su tono claramente casuístico, nos abre la posibilidad de mentir por necesidad; pero en la *Metafísica*, en cambio, Kant se niega rotundamente esta posibilidad y favorece una interpretación rigorista.

En este segundo capítulo, mi objetivo es justificar la interpretación rigorista de la filosofía kantiana

apelando a la *Fundamentación de la metafísica de las costumbres*. Considero que, además de la evidencia dada en el primer capítulo, a partir de este último texto de Kant es posible extraer razones filosóficas y conceptualmente relevantes para sostener que la filosofía kantiana, al menos para el caso de la mentira, exige el cumplimiento irrestricto del deber de la veracidad. Todo mi argumento se basa en que si se examina la máxima de mentirle al asesino a la luz del imperativo categórico, esta resulta moralmente impermisible. Esta conclusión requiere, por supuesto, explicar la naturaleza y función del imperativo categórico. Para llevar a cabo esta tarea es necesario analizar previamente la distinción entre acciones que se llevan a cabo por deber y aquellas que solamente son conformes al deber (§2.1). A partir de aquí podré explicar cómo opera el imperativo categórico en sus distintas formulaciones (§2.2). Posteriormente, me detengo en la primera formulación del imperativo y muestro que dadas las características de la noción de un deber perfecto no es permisible mentirle al asesino (§2.3 y §2.4). Más adelante examino la segunda formulación del imperativo y concluyo que toda clase de mentira resulta contraria a esta formulación (§2.5). En ambas formulaciones del imperativo resulta moralmente impermisible mentir. Finalmente, discuto con Christine Korsgaard y respondo a una posible crítica a mi interpretación (§2.6). Con este análisis concluyo que la filosofía moral kantiana es rigorista con respecto al deber de la veracidad.

2.1 Acción por deber y acción conforme al deber

2.1.1 Buena voluntad y deber

La *Fundamentación de la Metafísica de las Costumbres* tiene como objetivo "establecer el principio supremo de la moralidad" (*F* 4:392). Esto significa que Kant no busca describir cómo es el comportamiento moral humano, y tampoco busca resolver problemas de aplicación de las normas morales. Su objetivo es distinto y –podríamos decir– más básico: establecer el fundamento de la moralidad. Lo que sí es cierto es que este fundamento permitiría, entre otras muchas cosas, orientar adecuadamente la vida moral, pues con él podríamos distinguir cuáles acciones son morales y cuáles no.

Kant inicia la *Fundamentación* afirmando que la única cosa buena sin restricción alguna en el mundo y fuera de él es una voluntad buena [*guter Wille*] (*F* 4:393). Esto no significa que no existan otros bienes en la vida humana. Por ejemplo, existen bienes propios del temperamento, como la inteligencia, el ingenio o el discernimiento; también existen talentos buenos y deseables, como el coraje o la perseverancia. E incluso hay bienes o dones de la fortuna como la riqueza y la salud. Pero todos estos bienes son deseables y buenos en un sentido instrumental, pues las disposiciones del temperamento, los talentos o la fortuna pueden ser usados también para fines malvados o criminales. En cambio, lo único que puede pensarse como intrínsecamente bueno es la buena voluntad.

La buena voluntad no obtiene su valor por lo que produzca o por su utilidad instrumental para conseguir determinado fin. Incluso, dice Kant, una buena voluntad que enfrente la adversidad de la fortuna o la carencia de otros bienes aun así "brillaría pese a todo por sí misma cual una joya, como algo que posee su pleno valor en sí mismo" (*F* 4:394). Esto no implica que la buena voluntad no pueda estar acompañada de utilidad (o de fracaso), lo relevante es que estas consideraciones instrumentales no afectan su valor intrínseco. Para seguir con la analogía, atender a la posible utilidad de la buena voluntad es como pulir la joya para comerciarla mejor, pero estos adornos no afectan en nada el valor en sí mismo de la buena voluntad.

Una característica esencial de la idea de buena voluntad es que se trata de un concepto en esencia prescriptivo. Cuando Kant afirma que no es posible pensar algo que sea bueno en sí mismo aparte de la buena voluntad, lo que está queriendo decir es que nuestra concepción de lo que es absolutamente bueno no se deriva de la experiencia, sino de aquello que podemos pensar o concebir. Puede que no exista en el mundo una instancia o prueba de buena voluntad, pero esto no significa que nuestra concepción de lo bueno cambie: aquello que es absolutamente bueno, independientemente de su existencia efectiva, es la buena voluntad.

Cuando comparamos esta noción prescriptiva de la buena voluntad con una descripción de la voluntad humana encontramos que existe una brecha, pues es un hecho que la voluntad humana

no es, por sí misma, una buena voluntad. No obstante, dice Kant, es la razón práctica la facultad encargada de cerrar esta brecha: "el auténtico destino de la razón tiene que consistir en generar una buena *voluntad buena en sí misma*" (*F* 4:396; énfasis en el original). De modo que la buena voluntad no es algo ajeno a nuestra naturaleza, pero tampoco es algo que se dé, digamos, naturalmente. Los seres humanos no somos buenos por naturaleza, pero, en tanto que seres racionales, tenemos la capacidad para cultivar una buena voluntad.

> El proceso de transformar una voluntad humana en una buena voluntad es un ejercicio voluntario y depende enteramente del agente. Sin embargo, este no es un proceso sencillo, porque nuestra naturaleza no se corresponde con las exigencias de la buena voluntad. De allí que la única manera de llegar a ser buenos —de llegar a ser agentes morales— es obligando, constriñendo o apremiando la voluntad para que llegue a ser buena. Solo a través del apremio es posible vencer las inclinaciones naturales de la voluntad18. "La importancia de la moralidad […] depende del hecho de que la voluntad humana es corrupta, y necesita auto-corrección racional y fortaleza de carácter" (Wood, 2008a, p. 29).

La constatación de esta brecha y su respectiva disolución explican la función y la importancia del concepto de *deber* [*Pflicht*] en la filosofía moral kantiana. Este concepto implica obligación [*Verbindlichkeit*], apremio [*Nötigung*[19]] o constreñimiento. En pocas palabras, para que la voluntad humana pueda llegar a ser buena necesita determinarse a través del apremio del deber. Este es el sentido de las palabras de Kant cuando dice:

---

[18] Curiosamente, Aristóteles tiene una imagen muy pertinente de esta idea: dado que tenemos ciertas inclinaciones que nos alejan de lo bueno, entonces "deberemos tirar de nosotros mismos en sentido contrario […] como hacen los que quieren enderezar las vigas torcidas" (*EN* 1109b 3-5).

[19] Es muy interesante notar que el sentido del concepto de apremio [*Nötigung*] pertenece al campo del Derecho. En términos técnicos, un apremio es un procedimiento judicial para hacer efectivo un mandamiento. Como se verá más adelante, esto explica, al menos intuitivamente, la relación entre el apremio y los imperativos.

> Para desarrollar el concepto de una buena voluntad que sea estimable por sí misma […], vamos a examinar antes el concepto de deber, el cual entraña la noción de una buena voluntad, si bien bajo ciertas restricciones y obstáculos objetivos que, lejos de ocultarlo o hacerlo irreconocible, más bien lo resaltan con más claridad gracias a ese contraste (F 4:397).20

Cuando Kant señala que "el concepto de deber implica el de buena voluntad y, en este sentido, lo contiene, quiere decir que en todas las acciones hechas por deber está presente una buena voluntad" (Rivera, 2014, p. 31). No obstante, las acciones por deber, dada la condición de la voluntad humana, están acompañadas de "restricciones y obstáculos subjetivos". Siguiendo a Faviola Rivera, un obstáculo subjetivo se contrasta con un obstáculo objetivo. Los obstáculos objetivos serían impedimentos en el mundo que impiden llevar a a cabo una acción, mientras que un obstáculo subjetivo "es algo en la disposición de la persona que interfiere con la realización de su propósito" (*ibid.*). Por ejemplo, la pereza, las inclinaciones o los sentimientos serían ejemplos de obstáculos subjetivos. Sucede entonces que en las acciones que hacemos por deber experimentamos este tipo de obstáculos subjetivos cuando se lleva a cabo la acción moral.

2.1.2 Acción por deber y acción conforme al deber

El interés de Kant en analizar las acciones por deber en las que experimentamos obstáculos subjetivos radica en que a partir de ellas se resalta con "más claridad" el concepto de deber (cf. Rivera, 2014, p. 32). Kant tipifica cuatro tipos de acciones: (i) aquellas contrarias al deber, (ii) aquellas conformes al deber que se realizan por una inclinación inmediata, (iii) acciones conformes al deber que se realizan con un propósito interesado, y (iii) acciones por deber. Kant dirá que solo las acciones por deber son genuinamente morales. No obstante, para entender la naturaleza de

[20] En la traducción de Faviola Rivera: el concepto de deber "*contiene* el de una voluntad buena, si bien bajo ciertas restricciones y obstáculos subjetivos, los cuales, sin embargo, sin que, ni mucho menos, lo oculten y hagan irreconocibles, más bien lo hacen resaltar por contraste y aparecer tanto más claramente" (*F* 4:397; énfasis añadido).

estas últimas, Kant contrasta las acciones por deber con aquellas que *no* son por deber (aquí caben las que son contrarias al deber o las que son conforme al deber). Al entender por qué una acción que no es por deber no es genuinamente moral, es más sencillo entender qué característica tiene una acción por deber.

Para empezar, las acciones contrarias al deber evidentemente son inmorales, así que no representan un caso interesante de análisis, "en ellas ni siquiera se plantea la cuestión de si pudieron haber sucedido por deber, dado que incluso lo contradicen" (*F* 4:397). En cambio las acciones conformes al deber [*pflichtmäßig*] que se realizan bien sea por interés o por inclinación inmediata tienen la apariencia de ser acciones genuinamente morales, pero no lo son. Como sabemos, las únicas acciones genuinamente morales son aquellas hechas por deber [*aus Pflicht*], y ellas no suponen inclinación inmediata ni interés. La pregunta ahora es cómo contrastar las acciones por deber de aquellas acciones que son conforme al deber que se realizan ya sea por inclinación inmediata o por interés. Kant examina cuatro situaciones diferentes para trazar este contraste:

1. El caso del tendero: este es un ejemplo de una acción conforme al deber por un propósito interesado. Imaginemos un tendero que no cobra de más a sus clientes inexpertos. Este tendero mantiene un precio fijo para todos, de modo que "hasta un niño puede comprar en su tienda tan bien como cualquier otro" (*F* 4:397). En este caso, diríamos, el cliente se ha visto servido honradamente. Sin embargo, dice Kant, esto no es suficiente para decir que el tendero haya obrado por deber, pues es posible que él haya actuado por interés propio: "su beneficio lo exigía" (*ibid.*). Es posible que el tendero haya tenido un interés en no perjudicar su reputación al cobrar de más a los clientes inexpertos; si bien sirvió honradamente, su acción no necesariamente fue realizada *por mor* del deber. En este caso se ve que la acción se adecúa a lo que el deber manda (servir honradamente), pero se ve también que es posible que en el tendero también esté presente un propósito egoísta e interesado para hacerlo: ganarse una buena reputación. Por tanto, concluye Kant, si el tendero ha actuado por interés, su acción no tiene genuino valor moral.

Este ejemplo muestra por qué una 'acción conforme al deber con un propósito interesado' no es

una acción por deber. El punto de Kant es que el comerciante no se "inclina a realizar *por sí mismo* el acto moralmente correcto de tratar con honradez a los compradores. El objeto directo de su inclinación […] es su propio provecho" (Rivera, 2014, p. 33; énfasis en el original). De esta manera, 'cobrar lo justo' es solo un medio para un propósito interesado. Y esta es la razón por la que este tipo de acciones (conforme al deber por un propósito interesado) no poseen valor moral genuino, las acciones genuinamente han de hacerse por sí mismas y no como medio para otro propósito.

¿Qué sucede con las acciones conforme al deber por inclinación inmediata? Este tipo de acciones presentan un caso más interesante porque "no es inusual que se suponga que ambas expresan una buena voluntad" (Rivera, 2014, p. 32). Además, en este tipo de acciones tampoco se presenta un propósito interesado, así que la acción no se hace como un medio para un propósito adicional. De manera que el contraste entre este tipo de acciones y las acciones por deber puede ser más difícil de trazar, pues pueden confundirse fácilmente. Kant presenta los otros tres casos:

2. El deber de conservar la propia vida: para Kant, conservar la vida es un deber y además es algo hacia lo cual estamos inclinados inmediatamente. Esta inclinación, sin embargo, empaña el valor moral de la persona que conserva su vida de esta manera: "el angustioso desvelo que tal cosa suele comportar para la mayoría de los hombres no posee ningún valor intrínseco y su máxima carece de contenido moral alguno" (F 4:398). La tesis de Kant es que una acción que se realiza por inclinación inmediata no es genuinamente moral. De hecho, Kant señala que si una persona padeciera infortunios y desgracias y deja de amar la vida, y aun así la conserva, entonces diremos que esta persona sí ha actuado por deber.[21] Su acción tiene valor moral porque no ha actuado por su inclinación inmediata, la motivación de su acción es cumplir con el deber. De manera que existe una diferencia entre las máximas cuando la acción es conforme al deber por inclinación inmediata ("conservo la vida por inclinación" o "conservo la vida por miedo") y la máxima cuando la acción

[21] "Cuando los infortunios y una pesadumbre desesperanzada han hecho desaparecer por entero el gusto hacia la vida, si el desdichado desea la muerte, más indignado con su destino que pusilánime o abatido, pero conserva su vida sin amarla, no por inclinación o por miedo, sino por deber, entonces alberga su máxima un contenido moral" (F 4:398).

es por deber ("conservo la vida por deber"). Aunque en todos los casos la persona conserve su vida, solo su acción conducida por la máxima por deber tiene genuino valor moral.

¿Por qué Kant niega valor moral a aquellas acciones cuyas máximas están configuradas por la inclinación inmedita? Siguiendo a Rivera, la respuesta a esta pregunta se ve en el tercer caso que presenta Kant.

3. El deber de ser caritativo: ser caritativo es un deber, pero —dice Kant— muchas veces viene acompañado de un "íntimo placer en esparcir júbilo a su alrededor y pueden regocijarse con ese contento ajeno en cuanto es obra suya" (*F* 4:398). En ese caso, la acción caritativa de la persona filántropa, por más que sea conforme al deber, no posee ningún valor genuinamente moral. Y si este filántropo, por cosas de la mala fortuna, se viera sumido en la desgracia y su ánimo se viera afectado al punto de que ya no sintiera compasión ni aprecio por los otros, y si aun así este hombre afectado cumpliera el deber de ser caritativo no ya por inclinación, sino incluso en contra de ella, entonces "y solo entonces posee tal acción su genuino valor moral" (*F* 4:398).

Al igual que sucede en el anterior caso, el filántropo que actúa por inclinación inmediata no realiza una acción con valor moral, y en cambio la acción del filántropo que ayuda al prójimo, incluso si ha perdido su inclinación a hacerlo, sí que tiene genuino valor moral. Aunque en ambos casos se ayuda al prójimo, solo el caso en el que no está presente la inclinación inmediata tiene genuino valor moral. No obstante, este caso reviste especial interés porque no es un caso de un motivo egoísta (como el ejemplo del tendero), ni tampoco es una acción cuyo objeto sea uno mismo (como la inclinación natural a conservar la propia vida). Además, en tanto que es una inclinación inmediata lo que motiva al filántropo, su acción tampoco es un medio para otro propósito, sino que se realiza por sí misma. ¿Por qué entonces esta acción conforme al deber no tiene genuino valor moral?

La respuesta de Kant es que esta acción, sin embargo, está acompañada de otras inclinaciones (Rivera, 2014, p. 37). Resulta que ser caritativo viene aparejado a la alabanza: ser caritativo resulta

"honorable y dign[o] de aliento y encomio" (*F* 4:398). El punto de Kant es que la inclinación inmediata de ayudar al prójimo "coincide por fortuna con lo que es conforme al deber" (F 4:398), y en esa medida es alabable, pero esto no es suficiente para otorgar genuino valor moral a la acción. "No es necesariamente malo actuar por inclinación, el punto es que es cuestión de suerte o fortuna que la inclinación coincida con el deber" (Rivera, 2014, p. 37). Si, por ejemplo, el filántropo que cae en la desgracia tuviera la inclinación natural hacia la indiferencia, su inclinación no estaría conforme al deber. Igualmente, quien conserva su vida por miedo o por amor a la vida, es solo cuestión de fortuna que estas inclinaciones coincidan con el deber, pero bien podría suceder que no hubiera tal coincidencia y que la persona abandonara la vida impulsado por su infelicidad.

La conclusión que extrae Faviola Rivera del análisis de los ejemplos 2 y 3 me parece muy acertada: dada la relación contingente entre la inclinación y la acción conforme al deber, la relación entre el acto moralmente correcto y el motivo por el que se hace ha de ser *necesaria* (Rivera, 2014, p. 38). Y dado que las inclinaciones nacen de nuestra condición natural sensible y el deber tiene su origen en la razón práctica, resulta que nada asegura una relación necesaria entre la inclinación y el deber.

4. El deber de ser feliz: finalmente, Kant presenta el cuarto ejemplo que, en mi opinión, es el más difícil de todos. Según el autor, todos los seres humanos tenemos "ya de suyo una poderosísima y ferviente inclinación hacia la felicidad, al quedar compendiadas en esta idea todas las inclinaciones" (*F* 4:399) [22]. Adicionalmente, los seres humanos tenemos el deber de "asegurar la

[22] La definición de felicidad presente en la *Fundamentación* difiere de la definición que da Kant en la primera *Crítica*: "La felicidad es la satisfacción de todas nuestras inclinaciones (tanto *extensive*, atendiendo a su variedad, como *intensive*, por lo que atañe a su grado, como también *protensive*, respecto a su duración)" (C1 A806, B834). Esta última definición compromete a Kant con la idea de que la felicidad es inalcanzable en esta vida, porque mientras vivamos es imposible satisfacer todas las inclinaciones y necesidades, dado que nuestra condición humana es ya una condición de necesidad constante. En cambio, en la *Fundamentación* la noción de felicidad tiene una función diferente; se la concibe como un deber indirecto porque la satisfacción de las necesidades es una condición indirecta para poder cumplir el deber moral. Si el ser humano no satisface sus necesidades se verá expuesto a actuar inmoralmente. De esta manera las necesidades e inclinaciones humanas resultan, de alguna manera, sirviendo a la moralidad. Un interesante estudio introductorio sobre el contraste de estas dos concepciones de felicidad se encuentra en Parra (2017, p. 43-

felicidad, pues el descontento con su propio estado, al verse uno apremiado por múltiples preocupaciones en medio de necesidades insatisfechas, se convierte con facilidad en una gran *tentación para transgredir los deberes*" (*F* 4:399; énfasis en el original). Una novedad del ejemplo es que el deber de la felicidad es indirecto, en la medida en que es una condición para cumplir los deberes morales, aunque no me detendré en este punto. Lo relevante para mi análisis es la función de este ejemplo para contrastarlo con las acciones por deber. Los ejemplos de quien conserva la vida a pesar de perder el gusto por ella, o del filántropo que ya no tiene inclinación a ayudar al prójimo, tienen en común que sus actos se llevan a cabo incluso cuando ya no hay una inclinación inmediata que coincida con el deber moral, incluso hasta puede que haya un deseo de actuar en contra del deber (p.e. en el caso de quien conserva la vida por deber, lo hace aun cuando "desea la muerte"). En el caso de la felicidad, como lo interpreto, hay un aspecto adicional: la acción que demanda el deber de la felicidad puede ser contraria a la inclinación que tenemos todos los seres humanos hacia la felicidad. Según Kant, "la prescripción de la felicidad […] causa un enorme perjuicio a ciertas inclinaciones" (*F* 4:399). La expresión "prescripción de la felicidad" ha de entenderse como el deber por la felicidad, y la felicidad es el "compendio donde se satisfacen todas las inclinaciones". El problema es que este concepto de felicidad no es preciso ni fiable, por tanto, la prescripción de la felicidad también hereda esta imprecisión y vaguedad. Por esta razón, dice Kant, no es sorprendente cómo "una única inclinación que es bien definida [precisa] con respecto a lo que promete y al momento en que puede ser obtenida su satisfacción, pueda prevalecer sobre una idea tan variable [como la prescripción de la felicidad]" (*F* 4:399). El ejemplo del enfermo de gota ilustra este punto: un enfermo de gota puede elegir comer aquí y ahora lo que le gusta, a pesar del dolor que esto le produzca. Su decisión se explica porque él calcula que el deber de la felciidad (que aquí están asociadas a conservar su salud) es una noción "infundada" sobre el futuro que puede resultar muy imprecisa y no fiable; en cambio si actúa siguiendo la inclinación inmediata (i.e. comer lo que le gusta) al menos "no aniquila el goce del momento" por unas expectativas tal vez inalcanzables.

52). Por otra parte, Faviola Rivera (2014) desarrolla de manera muy clara la relación entre felicidad y moralidad en su libro *Virtud, felicidad y religión en la filosofía moral de Kant.*

Es por esto que una única inclinación bien definida y concreta puede superar y prevalecer por encima de una idea tan imprecisa y vaga como la de la prescripción de la felicidad. Y concluye Kant, incluso en este caso, si este paciente no tuviera una inclinación hacia la felicidad, es decir, si no tuviera una inclinación a satisfacer sus inclinaciones (si, por ejemplo, no deseara comer lo que le gusta), con todo, esta persona "todavía tene una ley, que es propiciar su felicidad, no por inclinación, sino por deber" (*F* 4:399).

Según mi interpretación, el caso del deber de la felicidad hace más patente el conflicto entre el deber y la inclinación. Si bien los seres humanos poseemos una inclinación a la felicidad y además tenemos el deber de la felicidad, no siempre coinciden ambas cosas. La coincidencia (contingente) entre inclinación y deber que veíamos en los casos 2 y 3 no es evidente en el caso de la felicidad. Lo que satisface la inclinación no siempre satisface el deber. A diferencia del deber de la felicidad, la inclinación es precisa y fiable, es por eso que una persona puede preferir seguir la inclinación y no el deber. Y en un escenario en el que la inclinación está dirigida a algo que contraviene el deber de la felicidad (p.e. el deseo de comer lo que gusta sobrepasa el deber de conservar la salud), si la persona actúa guiado por el deber, entonces, y solo entonces podríamos estar seguros de que la persona actúa por deber.

> Nosotros podemos (y debemos) actuar por deber solo cuando ninguna razón de auto-interés o una inclinación empírica es suficiente para motivarnos a realizar la acción. Una acción puede ser hecha por deber, por tanto, solo cuando no hay tal inclinación empírica —frecuentemente, aunque no siempre, cuando algunas inclinaciones nos llevan a actuar en contra de la acción correcta. (Wood, 2008a, p. 26)

2.1.3 Sobredeterminación motivacional [*motivational overdetermination*]:

El análisis anterior sugiere que una acción por deber solo es posible cuando no hay ninguna inclinación ni interés presente. Esta tesis tiene consecuencias importantes para la comprensión de la filosofía práctica kantiana. Básicamente, nos enfrentamos a la pregunta de si una acción

sobredeterminada motivacionalmente tiene valor moral. La sobredeterminación motivacional es un término que hace referencia al fenómeno en el que el deber moral y la inclinación están dirigidos hacia la misma acción (Wood, 2008a, p. 26). El problema con la sobredeterminación motivacional es que impide reconocer cuál es el motivo real de la acción del agente. Este fenómeno es se puede ilustrar con el caso del tendero. Como hemos dicho, este es un caso de una acción conforme al deber en el que hay también un propósito interesado. Si el tendero no cobra de más a sus clientes no es precisamente porque el deber lo exige sino porque está obrando de acuerdo a su propio beneficio. ¿Cuál es entonces el motivo de la acción del tendero? La sobredeterminación motivacional nos impide saberlo

El fenómeno de la sobredeterminación es relevante para trazar la distinción entre acciones por deber y acciones conforme al deber porque, como he intentado sugerir en la sección anterior, Kant considera que los casos que podrían ser catalogados de sobredeterminación no cuentan como acciones por deber. Para este autor, los casos en los que hay un deber mezclado con una inclinación o con un interés no se cumple la condición propia de una acción llevada a cabo por deber. En el ejemplo del tendero, al menos como Kant lo presenta, la acción honrada no tiene lugar por deber porque aquel tenía un propósito interesado (*F* 4:397). Y lo mismo puede decirse respecto a los otros ejemplos de Kant. En el caso de conservar la vida, todos tenemos el deber de conservarla, pero también tenemos una inclinación inmediata y natural hacia ello. En el caso de la ayuda al prójimo, el deber de ayudar puede mezclarse con el "íntimo placer de esparcir júbilo" hacia el prójimo. Asimismo, el deber de asegurar la felicidad también puede coincidir con la "ferviente inclinación hacia la felicidad". En todos estos casos es posible que el deber y la inclinación estén dirigidos hacia la misma acción, y, por lo tanto, se presentaría el fenómeno de la sobredeterminación motivacional.

En todo caso, Kant no está interesado en encontrar el motivo real en los casos de sobredeterminación motivacional, pues su objetivo en los ejemplos que presenta no es rastrear los motivos que puede tener alguien a la hora de actuar; su objetivo, como ya he indicado, es examinar las acciones sobredeterminadas motivacionalmente para contrastarlas con aquellas realizadas por

deber. Solo aquellas acciones donde no hay sobredeterminación son candidatas a contar como acciones por deber. El interés de Kant está centrado en aquellas acciones en las que el agente está exclusivamente constreñido (internamente) por el deber, es decir, en las que este reconoce y acepta la validez objetiva de la ley. En la medida en que no exista este elemento de constricción, el agente no está actuando por deber (*cf.* Wood, 2008a, p. 26).

El concepto de sobredeterminación motivacional deja bastante clara la diferencia entre acciones por deber y acciones conformes al deber. Vimos que una acción por deber se lleva a cabo sin la presencia de ningún propósito interesado ni ninguna inclinación inmediata, y solo cuando la acción se realiza por el deber mismo es que ella adquiere genuino valor moral. Sin embargo, esto no significa que las acciones conforme al deber sean inmorales. De hecho, es posible que las inclinaciones que mueven las acciones conforme al deber no tengan "nada moralmente desacreditable" (Wood, 2008a, p. 27): no hay nada moralmente malo en la inclinación natural a conservar la vida, o en la inclinación a sentir satisfacción por ayudar a otros. De manera que las acciones conforme al deber incluso pueden ser "honorables y dignas de aliento y encomio" (*F* 4:398), como el mismo Kant reconoce. El punto es que el universo de la moral es más exclusivo y estrecho que el universo de lo que es encomiable.

2.1.4 Imperativo como deber-ser

Queda por explicar ahora qué es una acción por deber. La respuesta a esta pregunta exige definir primero los términos. ¿Qué está entendiendo Kant por deber? "El deber significa que una acción es necesaria por respeto hacia la ley" (*F* 4:400). "Ley", a su vez, significa un principio práctico de la razón que es objetivamente o universalmente válido para todos los seres racionales (*F* 4:421). Y por "necesidad", Kant se está refiriendo a una exigencia práctica [*Nötigung*], es decir, un tipo de constreñimiento (Wood, 2008, p. 26). Ahora bien, el tipo de constreñimiento que impone el deber es interno, no externo; se trata de un "ejercicio sobre uno mismo" (*ibid.*) de respeto a los principios correctos a la luz de la razón práctica.

Los seres humanos tenemos una naturaleza no enteramente racional. Esto significa que la razón no determina suficientemente a nuestra voluntad, porque ella se ve sometida por las inclinaciones y por condiciones subjetivas (*F* 4:413). Dada esta falta de coordinación entre la razón objetiva y las inclinaciones subjetivas se hace necesario apremiar a la voluntad humana, es decir, se hace necesario determinar esta voluntad que no es enteramente buena, de tal modo que ella que actúe según la representación de leyes objetivamente necesarias. En palabras más sencillas, dado que somos seres no enteramente racionales, debemos imponer sobre nuestra voluntad los dictados del deber.

El modo en el que apremiamos a la voluntad, según Kant, es valiéndose de un mandato que se expresa a través de imperativos. "La representación de un principio objetivo, en tanto que resulta apremiante para una voluntad, se llama un mandato (de la razón) y la fórmula del mismo se denomina imperativo" (*F* 4:413). Los imperativos, en tanto que apremian a la voluntad, son expresados mediante un deber-ser (*F* 4:413). De esta manera se evidencia la relación entre una ley objetiva de la razón y una voluntad no enteramente buena: la forma en la que una voluntad humana ha de actuar movida por leyes objetivas es a través de mandatos (esto es, imperativos) de la razón que determinan lo que debe-ser la acción humana. Este deber-ser no ejerce influjo sobre la voluntad por medio de la sensación o de inclinaciones meramente subjetivas.

Dicho sea de paso, para una voluntad absolutamente buena, es decir, para una voluntad que nunca se ve movida por inclinaciones, no es necesario ningún apremio porque la modalidad subjetiva de esta voluntad absolutamente buena "solo puede verse determinada por la representación del bien. De ahí que para la voluntad divina y en general para una voluntad santa no valga imperativo alguno: el *deber-ser* no viene aquí al caso, porque el *querer* coincide ya de suyo necesariamente con la ley" (*F* 4:413; énfasis en el original). Dado que los seres humanos no tenemos voluntades santas ni divinas, necesitamos imponer a nuestra voluntad imperfecta leyes objetivas.

2.2 Imperativo categórico

La forma en la que apremiamos a nuestra voluntad para que se ajuste a las leyes objetivas no puede ser más que a través de imperativos, es decir, de mandatos que establezcan lo que debemos hacer. De esta manera se explica que los principios prácticos, a través de los cuales se expresa el apremio, tomen la forma de imperativos:

> Todos los imperativos mandan *hipotética* o *categóricamente.* Los imperativos hipotéticos representan la necesidad práctica de una acción como medio para conseguir otra cosa que se quiere. El imperativo categórico sería el que representaría una acción como objetivamente necesaria por sí misma, sin referencia a ningún otro fin (*F* 4:414; énfasis en el orginal).

La función de los imperativos es expresar qué acción es buena, ya sea como *medio* para algo o buena en *sí misma.* Podemos entender esta distinción de la siguiente manera: cuando un agente determina un fin, está necesariamente aceptando un principio normativo que le exige realizar las acciones necesarias como medios para alcanzar dicho fin. Este principio es denominado *imperativo hipotético*. "Se llama 'imperativo' porque es una exigencia de la razón que requiere que el agente haga algo; es 'hipotético' porque esta exigencia gobierna nuestra acción solo bajo la condición de que queremos el fin en cuestión" (Wood, 2008a, p. 67; énfasis añadido). Los imperativos hipotéticos, sabemos, se contrasten con los *imperativos categóricos*. Estos últimos no tienen ninguna condición ni están determinados a un fin ulterior. "Los imperativos categóricos son categóricos porque su validez no está condicionada por un fin *previo*" (*ibid.;* énfasis en el orginal).

> Hay un imperativo que, sin colocar como condición del fundamento ningún otro propósito a conseguir mediante cierto proceder, manda este proceder inmediatamente. Este imperativo es *categórico.* […] Este imperativo puede ser llamado el de la *moralidad* (*F* 4:416; énfasis en el orginal).

Si la moralidad se expresa a través de imperativos categóricos, ello se debe a que la función del imperativo moral no es alcanzar un fin ulterior, sino apremiarnos a actuar independientemente de

nuestros deseos o fines personales (Wood, 2008a, p. 68). Así descrita, la naturaleza del imperativo categórico es su carácter incondicional: "el supremo principio de la moralidad no admite condiciones o excepciones, por supuesto, porque no hay nada superior en referencia a lo cual las condiciones o excepciones puedan ser justificadas" (*ibid.*). Los mandatos o leyes de la moralidad comportan una necesidad incondicionada y, por lo tanto, son válidos universalmente, es decir, son mandatos que hay que obedecer, incluso en contra de la inclinación.

Teniendo en cuenta la anterior, es posible ahora a examinar y caracterizar el imperativo categórico. Siguiendo a Wood (2008a, p. 66-7), contamos con el siguiente esquema del imperativo categórico:

Primera formulación:

(FLU) Fórmula de la ley universal: "Obra sólo según aquella máxima por la cual puedas querer que al mismo tiempo se convierta en una ley universal" (*F* 4:421; *cf. F* 4:402).

Con su variante:

(FLN) Fórmula de la ley de la naturaleza: "obra como si la máxima de tu acción pudiera convertirse por tu voluntad en una ley universal de la naturaleza" (*F* 4:421; *cf. F* 4:436).

Segunda formulación:

(FH) Fórmula de la humanidad como fin en sí misma: "Obra de tal modo que uses a la humanidad, tanto en tu persona como en la persona de cualquier otro, siempre al mismo tiempo como fin y nunca simplemente como medio" (*F* 4:429).

Tercera formulación:

(FA) Fórmula de autonomía: "[…] la idea de la voluntad de cualquier ser racional como una voluntad que legisla universalmente" (*F* 4:431). O "No elegir sino de tal modo que las máximas de su elección estén simultáneamente comprendidas en el mismo querer como

ley universal" (*F* 4:440).

Con su variante:

(FRF) Fórmula de reino de los fines*:* "Obra según máximas de un miembro que legisla universalmente para un reino de los fines simplemente posible" (*F* 4:439).

Kant propone estas formulaciones y las variaciones como derivaciones del concepto del imperativo categórico. Hay interpretaciones que indican que todas estas son formulaciones del mismo imperativo (Korsgaard, 1996), y también están quienes sostienen que cada formulación es diferente y que hay un proceso de refinamiento que va desde una versión muy básica (FLU) hasta una formulación más acabada y completa (FRF) (Wood, 2008a). No me interesa abordar esta discusión. Me interesa, más bien, ofrecer una reflexión sobre la función del imperativo categórico en la filosofía práctica de Kant. Según mi interpretación, la función del imperativo categórico es la de indicar qué máximas de acción son morales. En otras palabras, el imperativo categórico sirve como criterio para distinguir las máximas de acción morales de las que no lo son.

2.2.1 El procedimiento del imperativo categórico

El imperativo categórico funciona como un criterio moral, en la medida en que nos ofrece un procedimiento para probar la permisibilidad/ impermisibilidad moral de máximas de acción. Concuerdo con la interpretación de Hill (1971), Rawls (1980) y O'Neill (1975) quienes consideran que imperativo categórico es un procedimiento de decisión para el razonamiento moral. En este sentido, el imperativo categórico puede leerse como un test de moralidad. Se supone entonces que cualquier máxima de acción puede ser evaluada bajo el procedimiento del imperativo categórico (PIC) y de esta manera se supone también que a través de este procedimiento se cuenta con un método para derivar los deberes morales o incluso para construir el contenido de toda la moralidad (Wood, 2008a, p. 70). De acuerdo con la interpretación tradicional, el PIC se ha asociado específicamente a la primera formulación del imperativo categórico (FLU), aunque desde luego las otras formulaciones también tienen su respectivo procedimiento (ver §2.5).

Me detendré a explicar cómo funciona el PIC según la fórmula de la ley universal. Según esta formulación, las máximas de acción que son permisibles moralmente han de cumplir las dos siguientes condiciones:

(i) Condición de concebibilidad: Si tu máxima de acción puede ser pensada o concebida sin contradicción como una ley universal (o como una ley de la naturaleza)[23].

(ii) Condición volitiva: Si tu máxima de acción puede ser *deseada* como una ley de la naturaleza.

Estas dos condiciones son pasos que una máxima de acción tendría que cumplir para ser permisible moralmente a la luz del test del imperativo categórico. La condición de concebibilidad verifica si podemos concebir (pensar) la máxima como una ley universal sin contradicción, es decir, "determinar si se presenta una 'contradicción del pensamiento' al intentar concebir la máxima de ese modo" (Rivera, 2014, p. 101). Por otra parte, si la máxima es concebible como ley universal sin contradicción, entonces habrá que verificar si cumple la segunda condición del test del imperativo categórico: la condición volitiva. Esta segunda condición busca "determinar si existe alguna 'contradicción de la voluntad' al *intentar querer* que la máxima sea una ley universal" (*ibid.*; énfasis añadido). Igual que sucede con la primera condición, si la máxima no aprueba la condición volitiva entonces no es permisible moralmente. Por otra parte, si puedo querer que la máxima sea una ley universal, es decir, si no se presenta una contradicción de la voluntad, entonces la máxima

[23] Faviola Rivera (2018) ofrece una interpretación novedosa sobre la condición de concebibilidad. Esta autora sostiene que la concebibilidad universal de una máxima ha de tomar en serio la analogía de que la ley universal debe ser pensada como *ley de la naturaleza,* y no como la posibilidad de que la máxima sea "aprobada por todos al mismo tiempo" (Rivera, 2018, p. 56). Según su interpretación, pensar una máxima como ley universal consiste en "concebirla como una ley según la cual cierto efecto se sigue necesariamente de cierto fundamento con base en el cual la voluntad se determina a actuar" (interpretación de la universalización como necesidad) (ibid, p. 57). Con esta interpretación, la exigencia de la condición de concebibilidad exige que "actuemos según máximas que podamos querer como leyes que determinen efectos de manera necesaria" (*ibid.* p. 72) (como si fueran leyes naturales). Esta interpretación es valiosa porque cuestiona el concepto de universalización en el contexto del imperativo categórico: ¿qué significa que una máxima pueda ser concebida universalmente como ley de la naturaleza?

es moralmente permisible. Como se ve, tanto la condición de concebibilidad como la condición volitiva verifican que la máxima examinada no de lugar a contradicciones ya sea del pensamiento o una contradicción de la voluntad.

Una vez claro en qué consiste la prueba del imperativo categórico, y con el fin de entender cómo opera el PIC en casos concretos, tomo los ejemplos que presenta Kant en la *Fundamentación*. Hay que aclarar, sin embargo, que existen interpretaciones distintas del PIC, interpretaciones que el mismo Kant no tiene en cuenta (Korsgaard, 1985, p. 43). Esto hace más difícil la tarea de entender la función del imperativo categórico. Más adelante explicaré las distintas interpretaciones del PIC, pero por ahora me interesa analizar los ejemplos siguiendo de cerca el texto de Kant.

1. Máxima del suicidio: Una persona que haya caído en desgracia y experimente un "hastío hacia la vida" puede preguntarse si acaso "no será contrario al deber para consigo mismo arrebatarse la vida" (*F* 4:422). Esa persona, entonces, se detiene a reflexionar si esta acción será o no contraria al deber moral. Su máxima de acción sería: basado en el egoísmo, el bienestar o el interés propio adopto el principio de quitarme la vida cuando esta me amenace a largo plazo con más desgracias que ventajas. En este punto la persona aplicaría el PIC: ¿Podría pensar sin contradicción que esta máxima fuera una ley universal de la naturaleza? Kant afirma que si la máxima no pudiera ser pensada por quien tiene la intención de suicidarse como ley natural, él estaría dando lugar a una contradicción: "se advierte que una naturaleza cuya ley fuera destruir la propia vida por esa misma sensación cuyo destino es impulsar el fomento de la vida se contradiría a sí misma y no podría subsistir como naturaleza" (*F* 4:422). En otras palabras, una persona que piense en suicidarse lo hace en nombre del bienestar o en nombre del interés propio, sus desgracias le hacen imposible vivir la vida que quiere vivir. Si esto es así, sería contradictorio que una persona elimine su vida justificada en el interés por aliviarla. En este sentido, la máxima del suicidio que al universalizarse resulta contradictoria no logra cumplir la condición de concebibilidad del PIC, por tanto, es una máxima de acción inmoral.

2. Máxima de la promesa falsa: Una persona en situación de necesidad pide dinero prestado sabiendo

que no va a devolverlo; sabe que debe mentir para obtener el préstamo. La máxima de acción de esta persona sería: "cuando me crea sumido en un apuro económico, pediré dinero a crédito y prometeré devolverlo, aunque sepa que nunca sucederá tal cosa" (*F* 4:422). El resultado de la aplicación del PIC a este caso es que no es posible pensar sin contradicción la promesa falsa, pues convertir la máxima en una ley universal "haría imposible la propia promesa […] dado que nadie creería lo que se le promete" (*F* 4:422). En otras palabras, si fuera una ley universal prometer falsamente, ya nadie creería en ninguna promesa, la institución misma de la promesa se vendría abajo si la promesa falsa resultara universalizada.

3. Máxima del cultivo de los propios talentos: Una persona tiene un talento que, cultivado, resultaría útil para él y para la humanidad. Sin embargo, prefiere no cultivarlo y en vez de ello se procura placer y diversión. La máxima de acción de esta persona es: "si tengo un talento y mis condiciones son cómodas y placenteras, me dedicaré a la diversión y al placer antes que a cultivar ese talento". Es cierto que esta máxima puede pensarse sin contradicción como una ley de la naturaleza, pues si todos "dejan enmohecer su talento y se dedican al goce" la naturaleza aún podría subsistir (*F* 4:423). En otras palabras, esta máxima de acción cumple con la condición de concebibilidad; sin embargo, no cumple la condición volitiva: a un ser racional le "resulta imposible *querer* que esto se convierta en una ley universal" (*F* 4:423; énfasis en el orginal), pues en tanto ser racional quiere que se desarrollen todas sus capacidades, "ya que le son dadas y resultan útiles para toda suerte de propósitos" (*F* 4:423). ¿Cómo pensar la posibilidad de que alguien quiera y no quiera al mismo tiempo desarrollar sus capacidades?

A diferencia de lo que sucede con el ejemplo de la promesa falsa o con el del suicidio, la contradicción que se presenta en el caso del cultivo de los talentos es de un tipo diferente. Se trata de una contradicción de la voluntad: la contradicción surge porque la voluntad quiere algo que no puede querer porque iría en contra de su misma constitución. Pensemos esta contradicción volitiva con el ejemplo de la máxima que favorece el goce y la pereza por encima del auto-perfeccionamiento. En este caso, dice Kant, habría una contradicción volitiva porque una voluntad humana busca su beneficio y utilidad, y esto lo conseguiría mejor desarrollando en alguna medida

sus habilidades; por tanto, sería contradictorio volitivamente que un ser racional con una voluntad humana desee desperdiciar sus habilidades, pues ello iría en contra de su tendencia volitiva al beneficio y la utilidad.

4. Máxima de la autosuficiencia: Una persona tiene buena fortuna económica, pero percibe que los otros tienen dificultades y necesidades materiales. Esta persona considera que la situación económica de los otros no es su problema, "¡Que cada cual sea tan dichoso como el cielo quiera o pueda hacerse a sí mismo!" (*F* 4:423). Siguiendo a Rivera, la formulación explícita de esta máxima sería: "cuando me vaya bien o no tenga ganas de hacerlo, no contribuiré en nada al bienestar ajeno" (Rivera, 2014, p. 125). Si esta máxima de acción fuera una ley de la naturaleza, el género humano podría subsistir. Esta máxima de acción cumple la condición de concebibilidad; pero, como sucede en el ejemplo anterior, es imposible querer que esta sea universal: la condición de la voluntad humana busca el interés propio y el beneficio, en algunos casos precisa compasión o ayuda por parte de los otros, y espera recibir ayuda si la necesita; pero si la máxima de acción de la autosuficiencia fuera una ley universal, entonces esta voluntad que puede en algún momento precisar ayuda eliminaría el auxilio que desea para sí misma.[24]

La clave para entender el PIC depende de qué signifique que una máxima moral se contradiga cuando se universaliza. He dicho que el PIC establece dos condiciones, la primera (condición de concebibilidad) verifica si la máxima de acción puede ser pensada o *concebida* sin contradicción

[24] Kant explica esta contradicción de la siguiente manera:

> Hacer el bien, es decir, ayudar a otros hombres necesitados a ser felices, según las propias capacidades y sin esperar nada a cambio, es un deber de todo hombre. Porque todo hombre que se encuentra necesitado desea que los demás hombres le ayuden. Pero si él manifestara su máxima de no querer, por contrapartida, prestar ayuda a los demás en caso de necesidad, es decir, si la convirtiera en una ley permisiva universal, todos le negarían igualmente su ayuda cuando él mismo estuviera necesitado o, al menos, tendrían derecho a negársela. Por tanto, la máxima del interés personal se contradice si se convierte en ley universal, es decir, es contraria al deber, por consiguiente, la máxima del interés común, consistente en hacer el bien a los necesitados, es un deber universal de los hombres […] (*MC* 453).

como ley universal; la segunda (condición volitiva) verifica si uno puede *querer* que la máxima de acción se convierta en ley universal. En otras palabras, la primera condición aborda una perspectiva objetiva (la máxima pensada como una ley universal válida para todos los seres racionales) y la segunda aborda una perspectiva subjetiva (la máxima como objeto posible del querer). Estas dos condiciones dan lugar a dos tipos distintos de contradicción tal y como puede apreciarse en los ejemplos analizados. Los ejemplos 1 y 2 ilustran una contradicción que viola la condición de concebibilidad, mientras que los ejemplos 3 y 4 muestran una contradicción que viola la condición volitiva.[25]

Ahora bien, si una máxima de acción no aprueba ninguna de las condiciones del PIC, entonces la máxima no es permisible moralmente, es una máxima inmoral. Esto ilustra la manera en la que el imperativo categórico funciona como un criterio o test moral: las máximas que aprueban el procedimiento son permisibles moralmente y por lo tanto está bien guiar la acción de acuerdo con ellas. Y si una máxima no aprueba el procedimiento, ello indica que el deber consiste en hacer lo contrario de lo que ella propone, "más bien debe permanecer como una ley [universal] lo contrario de dicha máxima" (*F* 4:424). Puesto en otros términos: dado que la máxima de prometer falsamente o la máxima de quitarse la vida no pueden ser pensadas ni queridas como leyes universales, entonces lo que debe ser una ley universal es el deber de conservar la vida y el deber de la veracidad, respectivamente.

¿Qué significa entonces que una máxima se contradiga cuando se universaliza? Para responder

[25] Kant categoriza estos ejemplos a través de la distinción entre deberes hacia uno mismo y deberes hacia los otros (*F* 4:421, 4:429, 4:430n). El ejemplo del suicidio y el ejemplo del auto-perfeccionamiento dan lugar a deberes hacia uno mismo: el deber de preservar la vida y el deber de desarrollar los talentos, respectivamente. En cambio, el ejemplo de la promesa falsa y el de la ayuda al prójimo dan lugar a deberes hacia los demás: el deber de la veracidad y el deber de benevolencia. Los deberes hacia uno mismo son aquellos que tenemos en virtud de honrar la condición humana propia, los deberes hacia los otros dependen de la condición de humanidad que tienen los otros seres racionales. Esta distinción, a su vez, será central en la *Metafísica de las costumbres*. Allí Kant expone toda la Doctrina Elemental a través de la distinción entre 'deberes del hombre hacia sí mismo' (Primera parte) y 'deberes éticos hacia los demás' (Segunda parte).

esta pregunta es necesario entender las distintas interpretaciones de cómo funciona el procedimiento del imperativo categórico en su primera formulación (Fórmula de la Ley Universal). Estas distintas interpretaciones están de acuerdo en que el objeto de evaluación del imperativo categórico es la máxima. También concuerdan en que esa máxima se universaliza y conduce a la pregunta de si el agente puede querer que su máxima de acción pueda ser pensada como ley universal. La máxima del agente falla este test moral si no puede ser pensada sin contradicción. Siguiendo la categorización de Korsgaard (1985), existen al menos dos interpretaciones diferentes del procedimiento del imperativo categórico para la primera formulación: la interpretación práctica y la interpretación lógica.[26]

Interpretación práctica del PIC: De acuerdo con la interpretación práctica, una máxima de acción fracasa en el PIC cuando al ser universalizada genera una contradicción en sentido *práctico*: "la máxima universalizada se contradice a sí misma cuando la eficacia de la acción como un método para alcanzar su propósito es subvertida por su práctica universal" (Korsgaard, 2011, p. 271). O en otras palabras, "la máxima universalizada frustra la habilidad original de la máxima para cumplir el propósito del agente" (Herman, 1993a, p. 137).

¿Cómo opera la interpretación práctica frente a máximas que no cumplen la condición de concebibilidad? Tomemos una vez más el ejemplo de la promesa falsa. Es un sinsentido práctico realizar promesas falsas en un mundo donde sea ley prometer falsamente, pues la práctica universal de la promesa falsa elimina las condiciones de confianza que hacen posible que las promesas falsas sean eficaces (*cf.* Herman, 1993a, p. 137). No es posible que el agente se adhiera a su máxima de prometer falsamente y al mismo tiempo acepte la forma universal de la máxima, pues ya nadie

[26] Korsgaard también menciona una tercera interpretación defendida por Paton (1947, p. 146-157): la interpretación *teleológica* del PIC. Aquí solo tendré en cuenta la interpretación práctica y la lógica, pues estoy de acuerdo con Barbara Herman (1993) en que la interpretación teleológica se compromete con las nociones de 'necesidad', 'armonía' y 'propósito necesario', las cuales son muy difíciles de defender y resultan muy problemáticas. Sin embargo, hay que reconocer que existe evidencia textual en Kant que apoya esta interpretación, especialmente en la *Fundamentación* (ver *F* 4:395, 4:421, 4:423; *MC* 429).

creería en las promesas y por lo tanto su promesa falsa no tendría el efecto deseado.

¿Cómo opera la interpretación práctica frente a máximas que no cumplen la condición volitiva? Una máxima no cumpliría la condición volitiva cuando, a pesar de que puede concebirse como una ley universal, entraría en conflicto con los propósitos que en tanto seres racionales debemos querer (Herman, 1993a, p. 137). Pensemos el caso de la máxima de la autosuficiencia (ejemplo 4): "Cuando me vaya bien o no tenga ganas de hacerlo, no contribuiré en nada al bienestar ajeno" (Rivera, 2014, p. 125). A pesar de que esta máxima cumple la condición de concebibilidad, ella no cumple la condición volitiva —en sentido práctico— porque los seres humanos esperan recibir ayuda cuando la necesiten. Aquí se presenta una contradicción de la voluntad: el propósito de la máxima de la autosuficiencia, cuando se piensa universalmente, se ve frustrado por el interés de los seres humanos por ser ayudados.[27]

Interpretación lógica del PIC: Esta interpretación es defendida por Dietrichson (1964), Kemp (1958) y Wood (1972). Según estos autores, una máxima falla el PIC cuando no puede ser concebida como una ley universal sin implicar una contradicción lógica. La interpretación es especialmente convincente en el caso de la promesa falsa; sus defensores tienden a enfocarse en la observación de que la promesa en sí misma sería imposible si se universalizara la máxima de la promesa falsa.

¿Cómo funciona la interpretación lógica frente a máximas que no cumplen la condición de concebibilidad? Es muy importante resaltar que la máxima de la promesa falsa no es contradictoria consigo misma. La contradicción se da entre la máxima y su forma universalizada: la máxima de la promesa falsa es contradictoria lógicamente con la versión universalizada de la promesa falsa. La contradicción tendría la forma p y no-p. La ley universal de la promesa falsa (p) haría imposible la máxima de la promesa falsa (no-p). La promesa falsa solo es posible si no es el caso que exista

[27] "[A]un cuando es posible que según aquella máxima pudiera sostenerse una ley universal de la naturaleza, es con todo imposible querer que un principio semejante valga por doquier como una ley natural. Pues una voluntad que decidiera eso se contradiría a sí misma, al poder darse algunos casos en que precise amor o compasión por parte de otros y en los que, merced a una ley natural emanada de su propia voluntad, se arrebataría la esperanza de auxilio que desea para sí" (*F* 4:424).

una ley universal de la promesa falsa, pues esta ley haría imposible la idea misma de promesa falsa: nadie podría prometer falsamente si no estuviera vigente la institución de la promesa. Esta interpretación muestra la tensión lógica entre una máxima inmoral y su forma universal. Literalmente, "no puedes concebir la promesa falsa como una ley universal porque nada puede ser una ley universal si no puede tener una instancia" (Herman, 1993a, p. 139). En otras palabras, la ley universal de la promesa falsa haría imposible (inconcebible, impensable) cualquier instancia de la máxima de la promesa falsa, pues si todos prometieran falsamente ya nadie creería en ninguna promesa, y por tanto ya no sería posible prometer falsamente.

¿Cómo funciona la interpretación lógica frente a máximas que no cumplen la condición volitiva? Según Herman, las máximas que no cumplen la condición volitiva fallan porque "cuando son deseadas como leyes universales contradicen algo que uno debería querer en la medida en que uno lo desea" (1993a, p. 137n6). Confieso que no logro comprender la interpretación lógica a la luz de la condición volitiva, pero, en términos generales, la contradicción consistiría en que uno no puede desear y no-desear lo mismo al mismo tiempo: mi deseo expresado en la máxima de prometer falsamente (que incluye ver realizado el objetivo para el cual miento) es contradictorio con mi deseo de que esta máxima sea una ley universal. Ambos deseos entran en contradicción porque desear la ley universal socava la posibilidad de cumplir el deseo particular de prometer falsamente (con éxito, cabría agregar).

2.3 Deberes perfectos y deberes imperfectos

Aclaradas las interpretaciones del PIC y la forma en la que funciona el imperativo categórico en su primera formulación, ya contamos con elementos suficientes para explicar la diferencia entre los deberes perfectos y los deberes imperfectos. Esta distinción básicamente depende de la diferencia entre la condición de concebibilidad y la condición volitiva. Me explico. Si una máxima de acción (p.e. prometer falsamente) no aprueba ninguna de los dos condiciones del PIC, entonces la máxima contraria (ser veraz) debe ser una ley universal en el sentido más estricto e ineludible, es decir, la máxima contraria es un deber perfecto. Por el contrario, si una máxima de acción

aprueba la condición de concebibilidad (la máxima se puede pensar sin contradicción como ley universal), pero no aprueba la condición volitiva (no se puede querer que la máxima sea una ley universal); entonces la máxima contraria es un deber meritorio o imperfecto. Por ejemplo, la máxima de no cultivar el talento puede ser pensada sin contradicción como ley universal, pero en tanto ser racional, un ser humano no puede querer que ella se convierta en ley universal. Es muy importante aclarar que si una máxima de acción no aprueba la condición de concebibilidad necesariamente no aprueba la segunda, pues si un ser racional no puede pensar una ley universal contradictoria, mucho menos podría querer que tal máxima se convirtiera en ley universal, ya que los seres racionales no pueden aceptar ningún tipo de contradicción. Por tanto, una máxima de acción contraria al deber perfecto, en estricto sentido, no cumple con ninguna de las condiciones del procedimiento. Por otro lado, un deber imperfecto sería aquel que cumple la condición de concebibilidad, esto es, la máxima puede pensarse sin contradicción como ley universal, pero no cumple la condición volitiva. En este caso, la contradicción no se da a nivel del pensamiento, sino a un nivel volitivo.

Lo que dice Kant al respecto de la distinción entre deberes perfectos/imperfectos es lo siguiente:

> Algunas acciones están constituidas de tal modo que su máxima no puede ser pensada sin contradicción como ley universal de la naturaleza y mucho menos que uno pueda querer que deba volverse tal. En otras no cabe detectar esa imposibilidad interna, pero sí resulta imposible querer que su máxima sea elevada a la universalidad de una ley natural, porque semejante voluntad entraría en contradicción consigo misma. Se advierte fácilmente que la primera contradice al deber más estricto (ineludible), la segunda al más lato (meritorio). (F 4:424)

Este pasaje de la *Fundamentación* da cuenta de la relación que existe entre el PIC y la distinción entre deberes perfectos e imperfectos. Adicionalmente, en la *Metafísica de las Costumbres* Kant introduce esta misma distinción, pero en vez de hablar de deberes perfectos e imperfectos se refiere a ellos como deberes estrictos y amplios, respectivamente. "Cuanto más amplio es el deber, más

imperfecta es la obligación del hombre de obrar" (*MC* 390). La distinción puede entenderse de la siguiente manera: "entre más amplio o lato sea el deber, menor obligatoriedad tendrá una acción particular que pueda cumplir al deber" (Hochberg, 1974, p. 256).[28] Así, por ejemplo, el deber imperfecto de cultivar los propios talentos no tiene un nivel de obligatoriedad alto para ninguna de las acciones particulares que conduzcan a cumplir este deber. La razón de esta laxitud normativa es que "hay un número de cursos de acción que podrían llevarse a cabo para cultivar los propios talentos" (Hochberg, 1974, p. 256). Para el caso de los deberes perfectos o estrictos, se podría decir que si el deber es estricto, la obligación de realizar una acción es completamente apremiante. En el caso del deber perfecto de no hacer falsas promesas la acción debida es muy concreta: no prometer falsamente.

Como vemos, la fuerza de la obligación entre un deber perfecto y uno imperfecto es diferente. Los primeros son absolutamente incondicionales, los segundos no. Es posible explicar esta diferencia en términos de la necesidad que está involucrada en cada caso. Una violación a un deber perfecto va en contra de una necesidad racional incontrovertible ya que es —o debería ser— absolutamente inconcebible, para cualquier ser racional, que una máxima de acción sea contradictoria o genere contradicción. Actuar motivado por una máxima inmoral atenta contra la necesidad racional y, por lo tanto, sería una acción irracional. En último término, la exigencia de incondicionalidad de los deberes perfectos descansa en la necesidad de preservar la naturaleza racional. En cambio, las máximas de acción que violan los deberes imperfectos no atentan contra una necesidad racional, pero sí atentan contra una necesidad volitiva. Esta necesidad volitiva, pienso, tiene que ver con la necesidad de unidad de la voluntad (ver capítulo 5). La voluntad que entra en contradicción consigo misma estaría escindida, enfrentada consigo misma. Una acción contraria al deber imperfecto no sería irracional, pero sí sería *self-defeating*[29] [auto-limitante o auto-derrotante].

[28] Korsgaard critica esta interpretación en la que se identifican los deberes perfectos/imperfectos y los deberes estrictos/amplios (1985, p. 28ss).

[29] El adjetivo *self-defeating* se traduce usualmente como 'contraproducente'. Sin embargo, esta traducción es muy imprecisa en un contexto filosófico. Una mejor traducción sería *auto-limitante* o *auto-derrotante*. Estas traducciones capturan la idea de que la contradicción volitiva tiene lugar cuando alguien se compromete

Siguiendo una intuición que extraigo de Korsgaard, la fuente de la normatividad de los deberes imperfectos descansa en la necesidad práctica de que la voluntad sea una cuando actúa, pues de otra manera la acción no sería atribuible a un mismo agente (como un todo) sino a una parte de él o a una fuerza que opera *sobre* o *en* él (*cf.* 2009, p. 18)[30].

Los deberes perfectos, entonces, comportan una obligatoriedad que no tienen los deberes imperfectos. Los primeros no pueden no llevarse a cabo y suponen una obligación concreta y directa, mientras que los deberes imperfectos pueden "aplazarse" o "postergarse" en la medida en que no hay un curso de acción definido o único que los satisfaga. El deber perfecto tiene especificada la acción que satisface el deber; el deber imperfecto —en cambio— puede cumplirse de distintas maneras, de ahí que el apremio a actuar en favor de él sea menor.

2.4 La máxima de la mentira al asesino

Todo el análisis anterior es relevante para responder, desde una perspectiva kantiana, al interrogante ético del asesino en la puerta. En el primer capítulo se había alcanzado la idea de que la mentira es una acción contraria al deber moral, pero la respuesta al interrogante del asesino en la puerta no era aún concluyente. En este último apartado me interesa mostrar que, gracias a la noción de deber perfecto, este interrogante tiene una respuesta: debemos ser veraces con el asesino que nos pregunta si nuestro amigo se encuentra refugiado en casa.

---

con una máxima que, al ser universalizada, termina por contradecir su voluntad. En la interpretación práctica del PIC, esto significa que el propósito expresado en la máxima se vería frustrado cuando la máxima es universalizada. En la interpretación lógica, significa que el deseo de cumplir la máxima se contradice con el deseo de que la máxima sea una ley universal. En ambos casos, repito, no es posible pensar la universalización de la máxima sin contradecir la voluntad del agente.

[30] "Una acción requiere un agente, alguien a quien atribuirle el movimiento en cuestión como su autor. Y también creo que es esencial al concepto de agencia que el agente esté unificado. Es decir: para considerar algún movimiento de mi mente o de mi cuerpo como *mi acción,* debo verlo como una expresión de mi yo como un todo, más que como el producto de alguna fuerza que trabaja *sobre* mí o *en* mí" (Korsgaard, 2009, p. 18; énfasis en el original).

Hasta aquí he profundizado especialmente de la máxima de la promesa falsa, y he mostrado que cuando esta máxima se universaliza no cumple con la condición de concebibilidad ni con la condición volita, y por tanto, según la propuesta kantiana, es un deber moral perfecto no mentir. La cuestión ahora es entender el alcance de este deber perfecto. ¿Debemos ser veraces siempre –en cualquier situación y circunstancia– o este deber solo aplica para el caso de las promesas falsas?

La interpretación estándar asume que el ejemplo de la promesa falsa es la prueba de que la mentira, cualquier mentira, es una violación del deber perfecto de la veracidad. Sin embargo, esta interpretación es errónea porque, si bien quien promete falsamente está mintiendo, ello no significa que cualquier mentira constituya un caso de promesa falsa. Para mostrar que cualquier mentira es siempre mala, sería necesario mostrar que ninguna máxima que implique mentir puede ser universalizable. Y esto es algo que al menos Kant no intentó demostrar (*cf.* Wood, 1972, p. 614). Luego es legítimo preguntarse si acaso mentirle al asesino en la puerta constituye realmente una violación a un deber perfecto. Después de todo, la situaciones de prometer falsamente y la situación de mentirle al asesino son dos muy diferentes, aun cuando las dos sean instancias de mentira. Para el caso que nos ocupa, la pregunta es si la máxima de mentirle al asesino aprueba el test del imperativo categórico. Mi respuesta a esta pregunta es que, si bien la máxima de mentirle al asesino es distinta a la máxima de la promesa falsa –y por tanto requeriría un análisis diferente del PIC –, el resultado es el mismo: tanto la máxima de la promesa falsa como la máxima de mentir al asesino no pueden ser concebidas como una ley universal ni tampoco pueden ser queridas como una ley universal. Para poder llegar a esta conclusión, sin embargo, es preciso recordar que el PIC puede ser interpretado en sentido práctico y en sentido lógico. La pregunta entonces es si la máxima de mentirle al asesino –"miente al asesino para salvar la vida de la persona refugiada en tu casa"– puede ser pensada como ley universal sin contradicción en sentido práctico y también en sentido lógico.

Interpretación práctica: la interpretación práctica del PIC se basa en la idea de que la máxima

universalizada frustra el propósito de la máxima particular. Ahora bien, si fuera una ley universal mentirle a los asesinos para salvar a la persona refugiada, entonces, cuando enfrente al asesino y le mienta, él no creería mi mentira. En un mundo donde todos mintieran para salvar a su amigo, ningún asesino creería la mentira, luego el propósito de mentir pierde eficacia. De esta manera, el propósito de mi máxima de mentir se vería frustrado, si mentirle al asesino fuera una ley universal. Si este argumento es correcto, bajo la interpretación práctica, la máxima de mentirle al asesino no aprueba la condición de concebibilidad. Por tanto, sigue siendo un deber perfecto ser veraz con el asesino.

Interpretación lógica: en el contexto de la interpretación lógica, la máxima de mentirle al asesino tampoco cumpliría con la condición de concebibilidad. Si fuera una ley universal mentirle al asesino, no habría ninguna instancia posible de mentira, porque el asesino sabría que lo engañamos y nuestro no-decir-la-verdad (nuestro intento por engañarlo) no contaría como una mentira, pues una de las condiciones constitutivas de la mentira es que la persona engañada no sepa que es engañada. La ley universal de mentirle al asesino hace imposible, en sentido lógico, la instancia de mentirle. Si todos mienten, nunca se podría engañar al asesino y, por lo tanto, nunca podríamos mentirle.

Tanto en la interpretación práctica, como en la lógica la máxima de mentirle al asesino no puede ser concebida como ley universal. Esto significa que es un deber perfecto no mentirle al asesino en la puerta. Esta conclusión, sin embargo, presupone que el asesino es explícito con sus intenciones criminales.

El escenario en el que el asesino explicite sus intenciones puede resultar improbable, aunque tampoco se trata de un escenario irreal. Basta recordar, por ejemplo, las redadas nazis donde los soldados alemanes irrumpían en las casas en busca de judíos ocultos (Hallie, 1979; Tarantino, 2009). En todo caso, también es posible otro escenario en el que el asesino esconda sus intenciones aunque, sin que él lo sepa, nosotros seamos conscientes de su verdadero propósito. ¿Acaso mentirle

al asesino que oculta sus intenciones cambia el resultado del PIC?

Ciertamente el caso del asesino que oculta sus intenciones es diferente a un caso en el que no lo hace, porque implica una situación de doble engaño: el asesino pretende engañar asumiendo que el interrogado no conoce sus intenciones criminales, y el interrogado miente asumiendo que el asesino desconoce que sus intenciones criminales son conocidas. El asesino cree que nuestra respuesta es verdadera –asume que no le mentiremos–, y el interrogado asume que él creerá nuestra respuesta. De esta manera se logra engañar al asesino, el interrogado saca provecho de la creencia errónea del asesino y también de su conocimiento de las intenciones criminales de este último.

En este nuevo escenario resulta que la máxima de 'mentir al asesino que oculta sus intenciones' (llamaré a esta máxima MO) puede ser concebida como ley universal, asumiendo, por supuesto, todo este intrincado juego de intenciones. Me explico: MO puede ser pensada universalmente porque el propósito de la máxima no se vería frustrado si existiera esta ley universal. Esto se explica porque el asesino cree que le decimos la verdad, y por lo tanto, cree que todavía están en juego las condiciones de confianza y veracidad que se establecen en la comunicación cotidiana (Williams, 2006d, p. 99ss). En contraste, estas condiciones se pierden, por ejemplo, con la máxima de la promesa falsa: si todos prometieran falsamente, la práctica misma de prometer perdería todo sentido porque ya nadie confiaría en ninguna promesa. El asesino pretende engañar haciendo pasar su pregunta como algo que es parte del contexto de la comunicación normal –donde se espera una respuesta veraz–, pero el interrogado sabe de su engaño y entiende que su pregunta es un engaño. Esta situación permite mentir sin que se deslegitime el supuesto fundamental de veracidad que soporta la comunicación, pues el asesino esperaría que la respuesta que recibe sea verdadera, aun cuando cualquier persona, en la misma situación, le mentiría.

Una condición para que MO funcione como ley universal es que el asesino no sepa que el interrogado es consciente de su intención criminal. Es más, si el asesino esconde su intención es

porque asume que el interrogado no la conoce –esto prueba, además, que el asesino cree que si el interrogado conociera sus verdaderas intenciones no le diría la verdad, con lo cual se muestra que el asesino asume que la norma (ley) es mentir a un asesino que explicita sus intenciones.

Tenemos entonces que la máxima de mentirle al asesino cuando explicita su propósito no puede ser pensada universalmente, y en cambio la máxima de mentirle cuando oculta su intención criminal sí puede ser universalizable[31]. En conclusión, bajo ciertas condiciones, es permisible mentirle al asesino en la puerta. Esta conclusión, evidentemente, es muy relevante por sus implicaciones con respecto al rigorismo kantiano. Si existe una posibilidad de mentirle al asesino, parece que la sombra del rigorismo se desvanece, después de todo MO aprueba el PIC.

Sin embargo, esta defensa no es satisfactoria. ¿Acaso la respuesta ética al interrogante al asesino depende del tipo de pregunta que haga el asesino y del tipo de información que posea el interrogado? Es cierto que cada máxima tiene características particulares y, por lo tanto, cada máxima puede dar lugar a resultados diferentes. No obstante, MO es una máxima bastante particular por el doble engaño que implica. "Esta máxima 'funciona' porque el asesino ha malinterpretado algo importante sobre su propia posición" (McCarty, 2012, p. 333) y además el interrogado es consciente de la ignorancia del asesino. Esto significa que la persona interrogada tiene una ventaja epistémica sobre el asesino. Cuando el interrogado le miente al asesino, está usando su ventaja para hacerle creer lo que le conviene. En otras palabras, está manipulando las creencias del asesino para obtener un resultado deseado. Y aunque el propósito de mentirle al asesino sea meritorio y valioso, lo cierto es que la razón práctica del asesino es usada como un medio para un fin circunstancial.

[31] McCarty (2012) también llega a la conclusión de que la máxima de mentirle al asesino es prohibida en un caso –cuando este explicita su intención– y permitida en otro caso –cuando esconde su propósito–. Este autor señala que un resultado como este es bastante contraintuitivo, pues significaría que los asesinos explicitarían siempre sus intenciones con el propósito de garantizar que recibirán respuestas verdaderas (*cf.* 2012, p. 332).

Varios filósofos se han detenido a pensar si la acción de mentir implica violar la autonomía de la persona engañada, incluyendo también la autonomía del asesino (Hill, 1991; McCarty, 2012; Schapiro, 2006). Incluso, en la misma línea del pensamiento de Benjamin Constant (§1.2), se ha discutido si el asesino puede ser tratado con la dignidad que merece una persona que sea miembro del Reino de los fines, o si sus intenciones criminales excusan que lo tratemos como un medio (Schapiro, 2006). Sin embargo, el argumento que ofrezco va en otra dirección. Lo que me interesa resaltar es que MO es una máxima que cumple la condición de concebibilidad, solo si a su vez se compromete con la manipulación del asesino. En otras palabras, el éxito de MO para aprobar el procedimiento del imperativo categórico en su primera formulación depende de que la persona interrogada aproveche su ventaja epistémica y trate al asesino como medio.

Si este argumento es correcto, resultaría muy extraño que el test del imperativo categórico en su primera formulación avale una máxima como MO que contradice abiertamente a la segunda formulación del imperativo categórico. Como vimos, la segunda formulación de este imperativo es la fórmula de la humanidad como fin en sí mismo (FH): "Obra de tal modo que uses a la humanidad, tanto en tu persona como en la persona de cualquier otro, siempre al mismo tiempo como fin y nunca simplemente como medio" (*F* 4:429). Y lo cierto es que la máxima de mentirle al asesino que oculta sus intenciones solo puede ser concebida como ley universal si se viola esta formulación. Este resultado desestima la validez moral de MO. Resulta muy sospechoso que las distintas formulaciones del imperativo categórico conduzcan a resultados contradictorios. La carga de la prueba queda del lado de los defensores de Kant: ¿cómo es posible que la validez universal de una máxima dependa de una relación instrumental entre personas?

2.5 La mentira: usar al otro como medio

La conclusión de la sección anterior es que la máxima particular de mentirle al asesino no aprueba el procedimiento del imperativo categórico, incluso en el caso especial de MO. Para alcanzar esta

conclusión asumí que el imperativo categórico tiene la función de derivar deberes –perfecto e imperfectos– y que el deber de la veracidad es exigible, al menos para el caso del asesino en la puerta. Como aclaré antes, el deber de no prometer falsamente no equivale al deber general de no mentir. De ahí que sea necesario tener en cuenta cada máxima particular. Lo interesante, como intenté mostrar, es que incluso atendiendo a la particularidad de la máximas, no es posible justificar o excusar la mentira contra el asesino. De manera que la sombra del rigorismo kantiano sigue intacta.

Hay una línea de argumentación diferente y que en mi opinión refuerza la conclusión anterior. El argumento de §2.4 solo atendía a la primera formulación del IC, con lo cual, insisto, se validaba la lectura rigorista de la filosofía moral kantiana. No obstante, en esta sección muestro que según la segunda formulación del imperativo categórico también está prohibido mentir, solo que esta prohibición abarca cualquier clase de mentira (Korsgaard, 2000b, p. 137). De esta manera, creo, se refuerza la conclusión alcanzada tras el análisis de la mentira a la luz de la primera formulación del imperativo categórico.

Cuando se aplica la segunda formulación del imperativo categórico al caso de la mentira, Kant sostiene lo siguiente:

> Por lo que atañe al deber necesario u obligatorio para con los demás, quien se propone hacer ante otros una promesa mendaz comprenderá en seguida que quiere servirse de algún otro hombre simplemente como medio, sin que dicho hombre implique al mismo tiempo un fin en sí. Pues es del todo imposible que aquel a quien quiero utilizar para mis propósitos mediante una promesa semejante pueda estar de acuerdo con mi modo de proceder hacia él, y, por lo tanto, resulta imposible que pueda albergar el fin de esta acción (F 4:429-30; énfasis en el original).

Siguiendo a Korsgaard, el pasaje de Kant usa las expresiones: "es imposible que esta persona pueda

estar de acuerdo con mi modo de proceder hacia él" y "le resulta imposible albergar el fin de esta acción". En otras palabras, la persona engañada no puede asentir ni al modo en que lo trata el mentiroso, ni al fin que este busca con su mentira. Según Korsgaard, estas dos son las condiciones que implica la fórmula de la humanidad para evaluar las máximas de acción (Korsgaard, 2000b, p. 138). Aunque se trata de dos condiciones distintas –una respecto al modo, otra respecto al fin–, me detendré en ellas como si fueran una única condición: la condición de asentimiento. Esta condición establece que si una máxima no permite el asentimiento de la otra persona –específicamente de la persona afectada o de la persona en quien recae la acción– entonces la máxima es contraria a la fórmula de la humanidad.

El asentimiento no se refiere a un asunto de gusto o de opinión, no refiere al hecho de que quien asiente considere apropiado asentir. Si esto fuera así, la evaluación moral de la máxima sería subjetiva. La condición es más fuerte: significa que, según sea el caso, es posible o imposible dar el consentimiento (Korsgaard, 2000b, p. 138). Por ejemplo, en el caso de la promesa falsa, la condición de asentimiento no se cumple porque la persona engañada literalmente no puede asentir, ella ni siquiera sabe qué le están ofreciendo. La mentira implica que la persona engañada no conozca el engaño, y por tanto, le resulta imposible asentir al procedimiento y al fin que tiene el mentiroso. La imposibilidad de alguien para asentir en el caso de la mentira pone a esta persona en una desventaja epistémica con respecto a la persona que miente. Quien engaña se está aprovechando de la ignorancia del otro y lo manipula para promover sus propósitos.[32]

> El punto es que cualquier acción que dependa por su naturaleza y para su eficacia en la ignorancia o en la impotencia falla el test [de la fórmula de la humanidad]. Mentir claramente cae en esta categoría de acción: esta solo engaña cuando el otro no sabe que es

[32] "De acuerdo con la Fórmula de la Humanidad, el engaño y la coerción son las formas más fundamentales de maltratar a otros —las raíces del mal. La coerción y el engaño violan las condiciones del posible asentimiento, y todas las acciones que dependen de su naturaleza y eficacia en su carácter coercitivo o engañador son tales que los otros no pueden siquiera asentir" (Korsgaard, 2000b, p. 140).

una mentira. (Korsgaard, 2000b, p. 139)

La interpretación de Korsgaard a la que me adhiero muestra que la mentira –y cualquier engaño, en general– es un tipo de acción que comporta un trato hacia otra persona como si ésta fuera un medio. Esto se explica más claramente porque una acción que prevenga o impida que la otra persona sea una "causa primera" es una forma de tratarla como medio, una cosa o una herramienta (Korsgaard, 2000b, p. 141). Y el engaño y la mentira, en efecto, tienen esta característica: alguien que no puede asentir no es claramente causa primera. Alguien víctima de la mentira o el engaño pierde control sobre sus creencias, no cuenta con una imagen realista de su situación, fracasa tanto en su acción eficiente en el mundo, como en su interacción con otros. El engaño y la mentira tratan a la persona engañada como un medio en un sentido muy específico:

> [T]rata la razón de la persona engañada como una causa mediata […] tu razón es usada como una máquina: quien engaña intenta determinar que palancas jalar para conseguir los resultados deseados a partir de ti. La mentira trata la *razón* de alguien como una herramienta". (Korsgaard, 2000b, p. 141)[33]

[33] Thomas Hill (1991) tiene una aproximación muy semejante, y relaciona la mentira con la pérdida de autonomía. Su análisis, sin embargo, se centra solo en el caso de las mentiras benevolentes, aquellas en las que se miente para fines morales o para impedir fines inmorales (por ejemplo, mentiras u omisiones a pacientes terminales o algunos casos de mentiras paternalistas). Según este autor, los seres humanos (adultos) tenemos el derecho a gobernar nuestras vidas del modo que elijamos. Este mismo punto lo expresa muy bien McCarty: "Para ejercer este derecho se necesita información precisa sobre sus circunstancias para tomar decisiones completamente racionales" (McCarty, 2012, p. 339). De esta manera, Hill considera que la mentira benevolente representa una violación a la autonomía de la persona.

> [Las mentiras benevolentes] privan a las personas de una imagen realista de su situación. En la medida en que tener esta imagen es necesario para un genuino control racional sobre la propia vida, las mentiras benevolentes fallan en promover este ideal de autonomía […] (Hill, 1991, p. 39). [Y agrega:] "la oportunidad para vivir en control racional de la propia vida se incrementa cuando hay personas que uno puede inequívocamente identificar como preparadas para ofrecer respuestas directas y honestas a preguntas precisas" (*ibid.*, p. 41).

En una palabra, mentirle al asesino es contrario a la primera y segunda formulación del imperativo categórico. Sin embargo, esta conclusión se alcanza en varios niveles y en distintas etapas. Cuando se atiende a la primera formulación del imperativo categórico –la formulación de la ley universal– es posible ver que la máxima de mentirle al asesino no aprueba la condición de concebibilidad. Si se reformula la máxima en el caso en el que el asesino oculta sus intenciones –MO–, parece que esta es concebible universalmente, pero el precio es que se atenta contra la fórmula de la humanidad, pues para universalizar la máxima habría que tratar al asesino como un medio. Por otra parte, cuando se atiende a la segunda fórmula del imperativo categórico –fórmula de la humanidad como fin en sí mismo–, también resulta que no es un deber ser veraz con el asesino. En este caso, sin embargo, el argumento tiene un alcance más amplio, pues todo tipo de engaño supone tratar al otro como una causa mediata y, por lo tanto, como medio para algún fin.

Con este análisis se responde al interrogante ético del asesino en la puerta: es un deber moral no mentirle al asesino en la puerta. Con esta afirmación, entonces, se ofrece un aspecto de la justificación de rigorismo que pesa sobre la filosofía moral kantiana.

2.6 Korsgaard: el derecho a mentir

Christine Korsgaard es consciente de las dificultades que genera el rigorismo kantiano, pues, según ella, éste implicaría que "nuestras obligaciones morales nos dejan indefensos frente al mal" (Korsgaard, 2011, p. 267). La teoría kantiana propone un ideal muy elevado de conducta y además nos exige cumplir con este ideal independientemente de las circunstancias, de las consecuencias, o de lo que otros agentes hagan o dejen de hacer. Como en el caso del asesino en la puerta, la obligación de veracidad trae consecuencias muy graves y contra-intuitivas, y aun así —dice Kant—

Dicho sea de paso, el caso del asesino en la puerta podría ser leído como una instancia especial de mentira benevolente, en la medida en que se busca evitar un crimen. Sin embargo, a diferencia de los casos típicos de mentira benevolente (a pacientes terminales o mentiras paternalistas), mentirle al asesino beneficia a un tercero (la persona refugiada) y no tanto al destinatario de la mentira (el asesino).

la "ley permanece con todo su vigor, porque manda categóricamente" (*F* 4:439).

Korsgaard trata de matizar la posición rigorista kantiana. Esta autora se compromete únicamente con la versión de la máxima en la que el asesino oculta sus intenciones, es decir, solo analiza el caso a partir de MO. Bajo esta condición, Korsgaard coincide con el análisis que he presentado en este capítulo. Esta autora considera que si se analiza MO bajo la primera formulación del imperativo categórico, es posible mostrar que mentir es permisible. Sin embargo, cuando MO es analizado desde la perspectiva de la segunda formulación de imperativo, la mentira resulta contraria al deber. En este sentido, el imperativo categórico ofrece resultados diferentes y contradictorios con respecto al caso del asesino en la puerta. Korsgaard no ve en esto un problema. Esta autora considera que la prohibición incondicional de la mentira "se deriva de un *ideal* atractivo de las relaciones humanas que es la base del sistema ético [kantiano]" (Korsgaard, 2011, p. 270; énfasis añadido). Si Kant está mal en su conclusión acerca de la mentira al asesino que llama a la puerta, "esto se debe a la interesante e importante razón de que a veces la moralidad misma permite o requiere algo que desde una perspectiva ideal está mal" (*ibid.*).

La respuesta de Korsgaard consiste entonces en proponer una teoría de doble nivel [*double level theory*]. Siguiendo la intuición de Rawls en *Teoría de la Justicia* (2006), Korsgaard cree que la filosofía moral debe dividirse en una teoría *ideal* y otra teoría *no-ideal*. El objetivo de la primera es determinar los principios que deben cumplirse en un estado de cosas en el que todos los agentes son racionales y en el que sus acciones están determinadas por el imperativo categórico. El objetivo de la segunda teoría es establecer principios aplicables en situaciones no ideales, es decir, en casos en los que los agentes, a pesar de ser racionales, actúan en contra de los mandatos morales. En otras palabras, Korsgaard, apoyada seguramente en el doble-aspectismo kantiano (ver §3.3), piensa que el resultado rigorista aplica solo desde una perspectiva ideal de la moralidad, pero no desde una perspectiva no-ideal. Para esta autora, desde una perspectiva no-ideal, las excepciones al deber de la veracidad son posibles. En este sentido, el caso del asesino en la puerta "no impugna la ética de Kant como un sistema ideal; antes bien, muestra que necesitamos de principios especiales para

lidiar con el mal" (Korsgaard, 2011, p. 270).

La propuesta de la teoría de doble nivel, dice Korsgaard, se podría ajustar a la teoría moral kantiana. Con respecto al tema de la mentira, Kant estaría hablando en un nivel ideal cuando plantea el deber incondicional de no mentir; pero esto no excluye la posibilidad de que haya una teoría no ideal en la que se permitan principios que den lugar a mentiras por necesidad o defensivas, como aquellos casos que el mismo Kant acepta en las *Lecciones de Ética* (ver §1.1).

Si Kant solo estuviera hablando a un nivel ideal, dice Korsgaard, se caería en un "formalismo vacío" (2011, p. 291). Los agentes deberíamos actuar siempre como si estuviéramos viviendo en un reino de fines, pero esto nos deja vulnerables frente al mal y nos vuelve indiferentes ante las consecuencias desafortunadas de nuestros actos. Introducir una teoría de doble nivel en la filosofía kantiana soluciona el rigorismo moral. Dice Korsgaard:

> Las Fórmulas de la Humanidad [segunda formulación del IC] y del Reino de los Fines [tercera formulación de IC] proveerán el ideal que gobierne nuestra conducta diaria. [Pero] al lidiar con circunstancias perversas puede ser que abandonemos ese ideal. En semejantes casos, podemos decir que la Fórmula de la Humanidad es inaplicable porque no está diseñada para usarse cuando lidiamos con el mal, pero puede no obstante seguir guiando nuestra conducta. Esta fórmula define la meta por la cual estamos trabajando y si nosotros podemos generar reglas de prioridad sabremos qué rasgos de la fórmula son los más importantes. La fórmula nos puede dar orientación acerca de cuál de las medidas que podamos tomar es la menos objetable. (Korsgaard, 2011, p. 296s)

Korsgaard termina sugiriendo que la teoría de dos niveles es un opción "disponible para el kantiano". La fórmula de la humanidad y la idea del reino de los fines serían formulaciones que establecen estados de cosas ideales, metas morales "a largo plazo para la humanidad" (Korsgaard, 2011, p. 300). Sin embargo, no vivimos en un mundo ideal, y por lo tanto, no siempre podemos actuar de acuerdo con el ideal moral. En estos casos, se corre el peligro de que por cumplir el ideal

de conducta nos convirtamos en una "herramienta del mal". "Sólo en circunstancias malvadas —y sólo entonces— el reino de los fines puede convertirse en un objetivo [regulativo] a buscar y no en un modelo de conducta a seguir: tener esto en mente nos proveerá de cierta orientación" (*ibid.*).

Korsgaard considera entonces que la tensión entre los resultados de la fórmula de la ley de la naturaleza y los de la fórmula de la humanidad se resuelve distinguiendo niveles. La fórmula de la ley universal funciona desde la perspectiva no ideal, para la cual es permisible mentirle al asesino; la fórmula de la humanidad, en contraste, operaría en el nivel ideal en donde sería inmoral la mentira. Esta estrategia funciona en la medida en que cada formula del imperativo categórico opera en un nivel distinto, cada uno tiene su campo de acción. Aunque ambas formulaciones den lugar a una contradicción, esta se resuelve porque son resultados que corresponden a niveles de aplicación distinta. Mentir resulta permisible en cierto contexto, pero no en otro. En la solución de Korsgaard, la permisibilidad de la mentira no implica abandonar la búsqueda del ideal regulativo.

Todo el argumento de esta autora descansa en lo que llamo el *supuesto de separabilidad*, es decir, la idea de que las distintas formulaciones del imperativo categórico no son contradictorios porque operan en niveles diferentes. Korsgaard, sin embargo, se equivoca al separar radicalmente el campo de acción de las dos formulaciones. En otras palabras, considero que el supuesto de separabilidad es incorrecto. Este supuesto se viene abajo cuando el resultado de una formulación supone la negación de la otra formulación del imperativo.[34] Me explico: tal y como Korsgaard plantea su idea, las dos formulaciones se oponen, pero no se contradicen, porque operan en categorías diferentes. Sin embargo, visto con atención, al menos en el contexto específico de la mentira al asesino, la formulación de la ley universal representa la negación de fórmula de la

[34] Más aún, el supuesto de separabilidad no es coherente con la idea de Kant de que las distintas formulaciones del imperativo categórico expresan la misma ley moral: "Las tres citadas maneras de representarse el principio de la moralidad sólo son en el fondo otras tantas fórmulas de una misma ley, cada una de las cuales incorpora dentro de sí a las otras dos" (*F* 4:436).

humanidad. Como ya lo mencioné en §2.4, MO –que es la versión de la máxima con la que se compromete Korsgaard– solo funciona si a la vez supone una manipulación del asesino. En este sentido, MO resulta permisible –concebible universalmente– solo si implica tratar a otra persona como medio y, por tanto, si niega la fórmula de la humanidad.

Como lo veo, la situación para Korsgaard es más problemática de lo que ella piensa: la permisibilidad de la mentira al asesino (MO) solo es posible negando el ideal regulativo de tratar a todas las personas como fines y no meramente como medios. La excepción por la que le mentimos al asesino equivale a una negación del ideal regulativo de entender a la humanidad como fin en sí misma. Para que la mentira sea permisible inevitablemente se tendría que tratar al asesino como un medio, se tendría que atentar necesariamente contra el ideal regulativo.

En este caso, proponer la teoría de los dos niveles conlleva asumir también el supuesto de separabilidad de las dos formulaciones del imperativo categórico. Pero este supuesto trae más problemas de los que soluciona, y, como suele decirse, se corre el riesgo de arrojar el agua sucia junto con el bebé: Korsgaard solo quiere evitar una contradicción entre los resultados de las dos formulaciones, pero termina comprometiéndose con algo más fuerte, pues la permisibilidad de MO en el nivel no-ideal supone la negación del nivel ideal. El supuesto de separabilidad tiene la función de evitar la interferencia entre las dos formulaciones del imperativo; pero esta solución no arroja este resultado, pues para que la mentira sea permisible en el nivel no-ideal, al menos en el caso del asesino en la puerta, habría que rechazar el nivel ideal. En este sentido, el supuesto de separabilidad resulta falso porque a pesar de que las formulaciones del imperativo funcionen a distintos niveles, todavía se mantiene la contradicción entre ellas.

Si no es posible sostener el supuesto de separabilidad, resulta que la mentira al asesino no deja de ser una acción inmoral, aun desde el punto de vista no-ideal, pues MO solo es permisible atentando

contra la fórmula de la humanidad[35]. MO solo cumple parcialmente las condiciones del imperativo categórico propias de la primera formulación –específicamente, cumple la condición de concebibilidad y la condición volitiva–; pero falla al cumplir las condiciones de la segunda formulación del imperativo –condición de asentimiento–. En este sentido, insisto, MO no aprueba el test del imperativo categórico, y por tanto no es una máxima moralmente permisible.

Como se puede apreciar, y a diferencia de lo que cree Korsgaard, las dos formulaciones del imperativo puestas a la luz de las implicaciones que la autora ha sugerido, se contradicen: actuar de acuerdo a una de ellas supone negar directamente la otra. ¿Cuál camino se ha de tomar? Korsgaard propone aceptar la permisibilidad de MO, ya que así el agente kantiano no quedaría "indefenso ante el mal". Además, si MO es permisible, la teoría moral kantiana podría responder a situaciones como las del asesino en la puerta. Según Korsgaard, podemos mentir para no ser usados para fines criminales y todo ello sin renunciar a la filosofía moral kantiana.

He tratado de mostrar que esta defensa tiene un alto precio: de cualquier modo la mentira sigue siendo inmoral. La permisibilidad ganada a través de la fórmula de la ley universal es un resultado obtenido por un juego de doble engaño que de una u otra forma cumple con las condiciones del procedimiento del imperativo categórico. Pero no puede ser que el éxito de MO dependa de que "las personas involucradas ignoren algo con respecto a su situación" (McCarty, 2012, p. 333). Sería extraño que la persmisibilidad moral de una máxima dependa de la ignorancia de las personas involucradas. Y en cambio, lo que sí es cierto es que MO atenta directamente contra la fórmula de la humanidad. ¿Acaso es preferible sacrificar la fórmula de la humanidad a cambio de la permisibilidad de una máxima tan sospechosa como MO?

[35] Michael Cholbi ha expresado una idea semejante: justificar una excusa para mentir no implica que la mentira deje de ser inmoral. En sus palabras: "Excusar un acto de engaño es sin embargo categorizarlo como un acto de engaño e implícitamente hay una referencia a la norma moral que desaprueba el engaño. Estar excusado del engaño todavía implica que se engaña" (2013, p. 444).

Para los intereses de este capítulo, esta crítica a la propuesta de Korsgaard va encaminada a mostrar que el rigorismo sigue en pie. Mentirle al asesino, es decir, aceptar la permisibilidad de MO pone la carga de la prueba en Korsgaard. Ella tendría que explicar al menos dos cosas. En primer lugar, tendría que explicar por qué un agente kantiano debería actuar en contra del deber moral. Segundo, y esto es algo que se desprende de lo anterior, tendría que explicar por qué este agente kantiano que miente al asesino habría de preferir guiar su acción de acuerdo a la primera formulación del imperativo y no de acuerdo a la formulación de la humanidad. Probablemente, y hablando especulativamente, estas respuestas revelen cierto nivel de incoherencia en el modelo de deliberación kantiana, aunque no me ocuparé de eso ahora. Lo cierto es que a la luz del análisis del imperativo categórico, la mentira al asesino resulta contraria al deber moral. Y esta conclusión es la que sustenta el cargo de rigorismo que pesa en contra de la filosofía mora kantiana.

II. Rigorismo moral kantiano y el problema de la integridad personal

3. Rigorismo kantiano

3.1 El rigorismo moral kantiano
3.2 Develando el rigorismo
3.2.1 El requisito prescriptivista
3.2.2 La imagen desencantada del mundo
3.3 El perspectivismo kantiano
3.4 La concepción del yo dividido
3.4.1 El concepto kantiano de libertad en la *Fundamentación de la metafísica de las costumbres*
3.4.2 La noción de libertad en la *Metafisca de las costumbres*: ¿cómo imputar acciones no morales?
3.4.3 La concepción del yo dividido

Como he tratado de mostrar en la primera parte, el tema de la mentira en la filosofía kantiana es indesligable de la pregunta por el cumplimiento incondicional de los deberes perfectos. Aunque esta investigación no se centrará en la mentira, considero importante partir de aquí porque a la base de la posición kantiana sobre la mentira subyace toda una concepción sobre el deber.

Como hemos vista, esta concepción del deber –específicamente, el deber perfecto– ha dado lugar a la imputación de que la filosofía kantiana es rigorista, ya que a la naturaleza de los deberes perfectos le es intrínseco su cumplimiento incondicional. El concepto de deber perfecto, aun cuando no es el único elemento de la filosofía kantiana, sí es muy significativo dentro de su propuesta deontológica: si hay algo que está realmente restringido hacer es aquello que vaya en contra del deber perfecto.

Dentro de los círculos kantianos no es bien recibida la imputación de rigorismo porque supondría que la filosofía práctica kantiana es radical y exigente en exceso. Estoy de acuerdo con los

defensores kantianos en ver el rigorismo como un problema. Sin embargo, a diferencia de muchos de los intentos kantianos por responder a esta imputación, creo que el rigorismo presenta un problema mucho más difícil de lo que las respuestas kantianas están dispuestas a aceptar. Esta segunda parte de la investigación tiene el objetivo, primero, de explicar en qué consiste el rigorismo kantiano (capítulo 3); y segundo, cuál es, en mi opinión, uno de los principales problemas de esta posición filosófica (capítulo 4).

3.1 El rigorismo moral kantiano

El rigorismo se define como el "exceso de severidad, principalmente en materias morales o disciplinarias" (RAE). Esta definición de diccionario atiende a nuestra comprensión común del término, pues cuando se dice que 'algo es riguroso' o 'exige rigor', lo que entendemos es que ha de cumplirse o llevarse a cabo irrestrictamente.

En un sentido más técnico y especializado, el concepto del rigorismo tiene historia tanto en la filosofía moral como en la teología. Desde la perspectiva de la filosofía mora, el rigorismo moral sostiene que todo acto moralmente relevante o todo agente moralmente responsable debe ser caracterizado como bueno o malo. Y esto supone un dualismo de la vida moral. No hay estados ni evaluaciones intermedias (Allison, 2012, p. 100). La confianza del rigorismo moral en discernir entre actos morales o inmorales o entre agentes buenos y malos presupone la existencia de criterios normativos muy claros y precisos para la evaluación moral.

En términos teológicos o religiosos, el rigorismo es una doctrina propia de la Iglesia Católica-romana que sostiene que en casos de incertidumbre o cuando hay dudas de conciencia sobre cuál es el mejor curso de acción, siempre debemos actuar siguiendo el deber o el mandato religioso. En el *New Catholic Dictionary* se define el rigorismo de la siguiente manera:

> La doctrina moral que asume que, cuando se presenta un conflicto entre dos opiniones,

> una en favor de la ley, y la otra en favor de la libertad, la ley debe ser observada siempre, incluso si la opinion en favor de la libertad es más probable o muy probable en comparación con la opuesta. (Rigorism, 1910)

Al igual que sucede con el rigorismo moral, el rigorismo religioso debe estar comprometido con un conjunto de normas o reglas que determinan cuál es el curso de acción correcto. El rigorismo religioso confía en que el seguimiento de las leyes o mandatos religiosos es suficiente para orientar la acción humana. Si bien la ley no atiende a *todos* los casos posibles, se confía en que la *generalidad* de la ley siempre ofrezca una respuesta o indique una dirección para guiar cualquier acción que pretenda ser correcta.

Al comparar estas dos aproximaciones encontramos que el rigorismo moral es una tesis sobre la *evaluación* moral de las acciones o del carácter de un agente, y por definición se compromete con un dualismo evaluativo: las acciones son buenas o son malas. En cambio, el rigorismo religioso es una tesis sobre la *aplicación* de los principios o deberes morales: en caso de duda, el mejor curso de acción es el que impone el mandato religioso. Estos dos tipos de rigorismo, en mi opinión, están vinculados estrechamente, pues para *aplicar* el deber es necesario *evaluar* o *identificar* previamente ese deber. En realidad estos dos tipos de rigorismo son las dos caras de una misma moneda. Mientras que el rigorismo moral hace énfasis en la evaluación moral, el rigorismo religioso enfatiza la aplicación de los deberes: para *actuar* de modo rigorista primero habría que *evaluar* de manera rigorista.

En el caso de una situación dudosa, por ejemplo, en el caso de un conflicto práctico, se supone que la persona que decide cuenta al menos con dos cursos de acción posibles. Según el rigorismo moral, solo un curso de acción es correcto y el otro, necesariamente, incorrecto. Y según el rigorismo religioso, en casos de conflicto, lo que debemos hacer es seguir el deber o mandato religioso. Con esto se explica por qué el resultado natural del rigorismo moral es el rigorismo en la práctica: si las acciones solo pueden ser buenas o malas, las máximas de acción morales se impondrán, sobre las

máximas de acción inmorales.

Si volvemos al planteamiento de Kant, nos encontramos con la imputación de que su filosofía práctica es rigorista. Con esto se quiere dar a entender que la teoría moral kantiana exige el cumplimiento estricto de los deberes morales. Así formulada, esta afirmación resulta imprecisa, aunque no es falsa. El rigorismo kantiano es una postura mucho más elaborada de lo que sugiere la anterior afirmación y requiere una explicación adicional.

El modelo kantiano de deliberación está basado en el imperativo categórico (§2.2). Y una de las consecuencias o implicaciones de este modelo es que la corrección o incorrección moral de una acción es intrínseca a la acción antes que extrínseca (Schapiro, 2006, p. 34)[36]. Esto significa que el valor moral de una máxima depende únicamente de las características y de la estructura propia de la máxima; no hace falta ningún otro criterio externo a la máxima para determinar su valor moral. Ese valor es independiente de las consecuencias específicas a las que esta acción conduzca y es también independiente de cualquier particularidad ajena a la máxima. En otras palabras, el valor moral de la máxima nada tiene que ver con consideraciones consecuencialistas o con características particulares del contexto. Las circunstancias externas no afectan en nada la calidad moral intrínseca de la máxima. Es por esta razón que en el caso de la mentira al asesino, por ejemplo, no importa cuál es la consecuencia de decir la verdad (previsiblemente, la consecuencia es la muerte de la persona refugiada en casa), ni tampoco importa el tipo de vínculo que se tenga con la persona refugiada (nada importa si se trata de mi amigo, de mi hermano o de un desconocido), ni siquiera es relevante si lo persiguen injustamente. A la luz del imperativo categórico solo importa si la máxima de mentirle al asesino cumple la condición de concebibilidad. Si esta máxima no cumple con las condiciones del imperativo moral, entonces se convierte en una máxima impermisible; más específicamente, si la máxima de la mentira al asesino no cumple con el imperativo categórico, entonces es un deber hacer lo contrario de lo que propone esa máxima (*F* 4:424), es decir, si la

[36] "Aquellos que insisten en que las acciones son o pueden ser correctas o incorrectas 'en sí mismas' e 'independientemente de sus consecuencias' son los deontologistas […]" (Isenberg, 1964, p. 463).

máxima de mentirle al asesino no aprueba el PIC, la veracidad al asesino es un deber perfecto.

El deber perfecto de la veracidad exige cumplimiento incondicional por encima de las consecuencias o de otras características propias del contexto. Y es a partir de la categoría de deber perfecto que, en mi opinión, se justifica el rigorismo kantiano. El rigorismo moral kantiano, con respecto a la verdad, es la tesis que sostiene que el deber perfecto de la veracidad ha de cumplirse incondicionalmente. De manera que frente a un conflicto práctico, como en el caso del asesino en la puerta, la mentira es inmoral sin importar cuáles sean las características específicas de la situación o cuáles sean las posibles consecuencias. Cuando el asesino nos pida información por el paradero de nuestro amigo, es nuestro deber decirle la verdad.

El rigorismo kantiano tiene una serie de características. En primer lugar, está comprometido con la incondicionalidad del deber. Esto equivale a decir que el cumplimiento o aplicación del deber no está —no debe estar— condicionado por ningún elemento externo al deber mismo. En este sentido, las consideraciones consecuencialistas, las motivaciones idiosincráticas del agente (Corbí, 2012) o las circunstancias contingentes del mundo (Nussbaum, 1995) son irrelevantes a la hora de acatar el deber.

En segundo lugar, y derivado de lo anterior, el rigorismo kantiano descarta la posibilidad de excepciones a la ley moral. La ley moral es absoluta, no hay lugar para derogaciones. No hay ningún fenómeno normativo que esté por encima de la ley moral. La normatividad del deber tiene una primacía práctica por encima de cualquier otro concepto práctico-moral (sobre este punto volveré en §6.2). Justamente esto es lo que explica por qué en caso de un conflicto práctico siempre ha de prevalecer el deber moral.

En tercer lugar, el rigorismo kantiano garantiza la universalidad de los deberes morales. La exigencia del deber perfecto es consistente y vinculante para todos los agentes racionales. Ningún agente racional podría actuar en contra de los deberes perfectos y salvarse de la acusación de estar

siendo irracional. Lo anterior evita que el deber moral sea impuesto arbitrariamente o que la moralidad sea aplicada discrecionalmente.

En cuarto lugar, y esto es lo que más me interesa resaltar, el rigorismo moral kantiano está comprometido con una comprensión muy precisa de la racionalidad práctica y de la deliberación moral. La racionalidad práctica kantiana traza una identidad entre razón y moralidad, lo cual significa que la moralidad de una acción está justificada racionalmente; y también significa que una acción inmoral está categorizada normativamente como irracional (Hoyos, 2014, cap.1). Esta identidad entre moralidad y racionalidad, por supuesto, implica que la deliberación moral ha de consistir en un proceso o procedimiento exclusivamente racional (ver §2.2). El sentido de la deliberación moral racional kantiana, entonces, es un procedimiento que evalúa la permisibilidad moral de las máximas de acción de un agente a la luz del principio moral del imperativo categórico. Lo que determina la acción no son los elementos particulares de la circunstancia y tampoco el contexto en el que se encuentre quien decide, ni siquiera importa quién sea la persona que lleva a cabo la acción. Todas estas consideraciones son abstraídas de la deliberación moral kantiana, y en nombre de una pretendida universalidad y una pureza de los principios, la norma debe ser impuesta. Las acciones particulares quedan entonces subordinadas al ideal regulativo del principio moral. Existe una primacía práctica del principio moral. Es este el que quisiera destacar como el rasgo propio del modelo de deliberación moral del rigorismo kantiano.

3.2 Develando el rigorismo

La posición de Kan sobre la mentira, específicamente su posición sobre el caso del asesino en la puerta, ha sido usada por los críticos como una *reducción al absurdo* de la filosofía práctica kantiana. La queja es que sería "insensata" una teoría moral que nos impele a decir la verdad en un caso como el del asesino en la puerta (Williams, 2006d, p. 110). Esta conclusión ha llevado a interpretar el rigorismo como una prueba de la insuficiencia de la teoría kantiana.

Al parecer el rigorismo es una posición tan débil, en términos morales, que no hace falta un argumento adicional para justificar esta debilidad. Es decir, un resultado tan contra-intuitivo como el del asesino en la puerta es suficiente para debilitar la teoría moral kantiana. ¡Qué sentido tiene criticar al rigorismo si se trata de una tesis tan evidentemente frágil! Concuerdo con los críticos de Kant en que el rigorismo es una posición inaceptable; sin embargo, echo de menos en ellos el intento por explicar cuál es el problema de esta tesis. Si bien el rechazo al rigorismo es una posición relativamente extendida, creo que hace falta ofrecer razones para explicar cuál es el problema de esta posición.

La estrategia argumentativa que seguiré con el fin de caracterizar el problema del rigorismo consiste en encontrar sus presupuestos internos y sus compromisos teóricos. Para usar la analogía de Günter Grass, se trata de ir "pelando la cebolla" hasta descubrir lo que está detrás o debajo del rigorismo kantiano. Con este propósito, expondré dos de los presupuestos teóricos del rigorismo kantiano (sección 3.2). Como espero mostrar, estos presupuestos a su vez están comprometidos con una metafísica (§3.3) y con una concepción de la agencia muy específicas (§3.4). En este capítulo, entonces, iré exponiendo cada uno de los presupuestos y compromisos del rigorismo kantiano que, en conjunto, forman lo que Williams llama la "psicología metafísica del moralismo kantiano" [*Metaphysical phsycology of kantian moralism*] (Pyle, 1994, p. 12). En el capítulo 4, criticaré estos compromisos y explicaré cuál es, en mi interpretación, el problema del rigorismo moral kantiano.

3.2.1 El requisito prescriptivista

Mi intuición fundamental contra el rigorismo moral kantiano es que nos puede obligar a hacer algo detestable. En el caso específico del asesino en la puerta, el rigorismo nos exige ser veraces, aun en contra de salvar la vida de la persona refugiada en casa. Enunciado de una manera más general, esto significa que los deberes perfectos en ocasiones demandan cursos de acción que generan en nosotros conflictos prácticos: hay situaciones en las que nos resistimos a actuar de

acuerdo al resultado de la deliberación moral. Esta queja no es nada nueva en la escena filosófica. De hecho, autores como Joseph Butler, en el s. XVIII, ya constataba que "no siempre hacemos lo que tras reflexionar haríamos, ni siquiera lo que tras reflexionar hemos decidido hacer: la reflexión no tiene un poder irresistible sobre nosotros"(citado por Korsgaard, 2000, p. 133; Butler, 2006, p. 61ss). Igualmente, Bernard Williams ha hecho no pocos esfuerzos por mostrar que cierta comprensión de la moralidad en muchas ocasiones atenta contra nuestros compromisos personales más profundos, al punto que en ocasiones acatar el deber moral significa dejar de ser uno mismo (1973; 1981; 2006a). Nomy Arpaly (2003), por su parte, ha argumentado muy convincentemente a favor de que el resultado de la deliberación racional no siempre es el mejor curso de acción moral. Estos autores tienen en común la idea de que la deliberación moralidad no es el único ni el más importante criterio para guiar la acción humana. En cuanto a mí, no puedo estar más de acuerdo con la perspectiva de estos autores y espero poder desarrollarla más adelante (capítulo 4 y 6).

La filosofía kantiana, sin embargo, tiene, al menos a primera vista, una forma de responder a estas críticas. Si la queja es que la moralidad no es el único criterio de corrección normativa, la respuesta kantiana es que la moralidad sí es el criterio más importante y, por tanto, cualquier otra consideración tendría que depender de ella o someterse a ella. En este sentido, se acepta que en ocasiones hay disonancias entre lo que el deber manda y otras consideraciones prácticas: conflictos entre la moralidad y nuestros propósitos de felicidad, o conflictos entre el deber y lo que es importante o significativo para nosotros. El punto relevante es que, desde la perspectiva kantiana, estos conflictos siempre han de resolverse a favor de la moralidad. Como en el caso del asesino en la puerta, el deber moral de la veracidad se impone por encima de cualquier otra consideración.

La propuesta kantiana no está muy interesada en considerar criterios prácticos distintos a aquellos derivados de su modelo de deliberación moral-racional. De hecho, esta es la expresión misma de la deontología kantiana, pues, por definición, el proyecto kantiano es una empresa prescriptiva; y parte de su objetivo es ordenar lo que se debe hacer independientemente de cómo sea el mundo,

independientemente de cualquier otro criterio con pretensiones morales que no sea racional.

Afirmar que la teoría kantiana es débil porque nos exige cumplir el deber por encima de nuestras capacidades, o por encima de nuestra naturaleza humana, no es una estrategia apropiada en contra de la filosofía moral kantiana. El mundo puede ser como sea, los seres humanos somos como somos, y eso no resta ninguna validez al hecho de que deba ser de otra manera. Es en esto en lo que consiste el requisito prescriptivista kantiano.[37]Los criterios prácticos no morales están anclados en la naturaleza humana y, por tanto, no tendrían por qué constituir un problema para una teoría de carácter prescriptivista. El requisito prescriptivista es un poderoso instrumento para acorazar el proyecto kantiano. No importa si nuestra naturaleza se adecua al deber moral, lo relevante es que, independientemente de cómo seamos, nuestras acciones deberían ser morales.

3.2.2 La imagen desencantada del mundo

El requisito prescriptivista solo puede funcionar si presupone a su vez una fuerte distinción entre 'lo que debe ser' y 'lo que es', ya que para establecer la caracterización de 'cómo *deben* ser las acciones humanas' se debe contar con una noción de 'cómo *son* las acciones humanas'. Dicho brevemente, el requisito prescriptivista presupone necesariamente la distinción entre el deber-ser y la naturaleza humana.

[37] "El deber debe ser una necesidad práctico-incondicionada de la acción; tiene que valer por lo tanto para todo ser racional (el único capaz de interpretar un imperativo) y *solo por ello* ha de ser también una ley para toda voluntad humana. En cambio, lo que se deduzca a partir de la peculiar disposición natural de la humanidad, cuanto se deduzca de ciertos sentimientos y propensiones e incluso, si cabe, de una particular orientación que fuese propia de la razón humana y no tuviera que valer necesariamente para la voluntad de todo ser racional, esto es algo que ciertamente puede proporcionarnos una máxima, mas no una ley, puede suministrarnos un principio subjetivo según el cual nos permitimos tener inclinación y propensión a obrar, pero no un principio objetivo conforme al cual quedemos *obligados* a obrar, aun cuando toda nuestra propensión, inclinación y orientación natural estuvieran en contra" (*F* 4:425; énfasis en el original).

La distinción entre deber-ser y la naturaleza humana no es un capricho conceptual. De hecho, ella tiene sentido dentro del contexto específico de una concepción muy precisa del mundo y del lugar que los seres humanos ocupan en él. Aquí sigo a Korsgaard cuando afirma que la concepción kantiana del mundo es el resultado de una "revolución" que marca la transición del mundo antiguo al mundo moderno (Korsgaard, 1996, p. 2ss). Me permito hacer un rodeo para explicar esta última idea. Una revolución conduce del mundo antiguo al mundo moderno. Esta revolución consiste en la forma en la que se ve el mundo. Pensemos por un momento en el mundo de Aristóteles. Para este filósofo el universo operaba teleológicamente. Las cosas en el mundo tendían naturalmente a un fin o *telos*. Las consecuencias éticas de esta idea metafísica son las que considero más interesantes. Para Aristóteles, la virtud o excelencia consiste precisamente en vivir de acuerdo con el fin natural humano. Lo que otorga valor ético a nuestro carácter y a nuestras acciones es que sean coherentes con, o mejor, que tiendan al fin natural humano, un fin ajustado a nuestra capacidad racional (EN 1097b28-1098a20). El valor ético del carácter de un ser humano (y sus correspondientes acciones) depende entonces de un criterio natural. La naturaleza misma es fuente de la normatividad.

En el mundo aristotélico existen la materia y la forma. La forma determina el fin de las cosas, y la materia idealmente se adecua a este fin. Esta misma idea, con algunos matices, se aplica también para los seres humanos, pues nuestras acciones han de adecuarse al fin humano que es esencialmente racional. En contraste, el mundo moderno carece de formas aristotélicas. Los descubrimientos científicos del s. XVI mostraron que la naturaleza seguía constantes y que era posible establecer leyes naturales con las mismas características de las leyes matemáticas (universales y necesarias). Estas leyes explican cómo funciona el mundo, lo hacen predecible y, por tanto, ya no es necesario pensar el mundo a partir de fines. Las leyes mecánicas reemplazan los fines. La revolución científica trajo la idea de que el mundo funciona como un mecanismo que opera no porque tienda a un fin, sino porque hay fuerzas (mecánicas) que generan e interactúan a su vez con otras fuerzas.

El movimiento mecánico, podría decirse, es intrínseco al mundo material. No es preciso proponer algo extrínseco a la materia —algo como la forma o el fin— para explicar su movimiento. Este movimiento no conduce a ningún lugar en particular, no tiene un fin preestablecido. Y, sin embargo, el movimiento mecánico sí puede predecirse y, por tanto, es controlable bajo la condición de que se conozcan las leyes naturales. La promesa de la ciencia moderna es que, si se descubren esas leyes, entonces es posible controlar la naturaleza. Así lo expresaba Bacon: "el conocimiento es poder". Si la naturaleza es controlable, entonces queda en poder del ser humano dirigirla hacia algún fin; pero esta vez se trata de un (algún) fin-humano, no del fin-natural aristotélico.

La visión del mundo kantiana es la de un mundo configurado por la ciencia newtoniana, un mundo de causas y efectos mecánicos que obedece a leyes determinables, cuantificables y medibles. En este mundo mecánico solo existen fuerzas que se hacen contrapeso entre sí, y cada una de ellas, a su vez, es el resultado de una cadena causal previa.

La principal consecuencia práctica de la concepción moderna del mundo es que en un mundo mecánico se corre el riesgo de perder el sentido de ser agente moral. Si todos los eventos del mundo están determinados por leyes naturales y si estos son explicados en términos de causas y efectos, entonces no hay lugar para un agente que pueda actuar con independencia de esta cadena causal. Y de paso, si no hay lugar para la agencia, evidentemente tampoco tendría sentido hablar de valoración moral, pues nadie podría ser valorado positiva o negativamente cuando sus acciones no dependen de él.

Una segunda consecuencia práctica de esta imagen del mundo es que si todo queda reducido a una relación de fuerzas mecánicas, no hay lugar a la idea de valor. El mundo moderno está despojado de valor. El mundo moderno es solo materia (fuerza y energía) que carece de formas y de valor. En efecto, esta metafísica trae consigo una "visión desencantada del mundo" (Corbí, 2012, p. 41).

La solución kantiana consiste en aceptar esta metafísica del mundo moderno sin negar la existencia del valor. Para Kant, el valor es *im-puesto* sobre el mundo. Y quienes tienen la tarea de llevar a cabo esta imposición somos los seres humanos. Según Korsgaard, Kant acepta la expulsión del valor intrínseco del mundo natural, y a la vez sostiene que el valor es una construcción humana (1996, p. 4). Así, la naturaleza es neutra valorativamente –el mundo es solo el conjunto de fuerzas y eventos causales– y nosotros, los seres humanos, proyectamos nuestros valores sobre él. La naturaleza entonces está detrás o debajo de las proyecciones humanas, y es así que el mundo se colorea valorativamente. Para Kant, los seres humanos hacemos parte del mundo natural y mecánico, y en esa medida estamos sometidos a la necesidad natural; pero eso no excluye la posibilidad de que seamos, también, seres valorativos, seres que tenemos y "usamos palabras normativas" (Korsgaard, 1996, p. 44).

¿Cómo logra el ser humano construir y proyectar el valor (moral) sobre el mundo mecánico natural? Kant considera que el valor moral se proyecta en el mundo a través de la forma de una ley práctica. Esta ley práctica tiene su origen en la razón. Y dado que esta es una facultad que tenemos todos los seres humanos, se sigue lógicamente que estas leyes sean universales y válidas para todo ser racional (que incluye a los seres humanos y también a los seres racionales no-humanos). Lo importante es entender que la razón no pertenece al mundo mecánico –ella lo trasciende. Este es el sentido en el que Kant considera que los seres humanos somos "pertenecientes al mundo inteligible" (*F* 4:457): tenemos una facultad que nos permite tomar distancia de la necesidad natural, pues somos conscientes de la "causalidad respecto de nuestras propias acciones" (*F* 4:449) (este punto será ampliado en §3.4).

Mientras que en el mundo aristotélico el criterio de normatividad son los fines naturales, en el mundo moderno este criterio ahora está constituido por la ley práctica. La diferencia es significativa: cuando el valor proviene intrínsecamente de la naturaleza, la fuerza del valor ejerce atracción sobre nosotros, porque es coherente con nuestra naturaleza; pero cuando el valor

proviene de la ley práctica, la fuerza del valor es la obligación [*Verbindlichkeit*], y en este sentido es una fuerza compulsiva (Korsgaard, 1996, p. 4). En efecto, la ley práctica viene dada en forma de *deber*: dado que la voluntad humana es imperfecta, es necesario apremiarla, forzarla, obligarla. El apremio [*Nötigung*] es siempre un contrapeso para nuestra naturaleza imperfecta y débil; es necesario obligarla para que se corresponda con el criterio normativo de la ley práctica. La forma en la que la ley se impone sobre la naturaleza es a través de la idea de deber, porque el deber supone necesariamente la idea de obligación. "El *concepto de deber* es ya en sí el concepto de una *coerción* (coacción) del arbitrio libre por la ley" (*MC* 379; énfasis en el original).

En tanto que seres que hacemos parte del mundo mecánico, los seres humanos inevitablemente estamos sometidos a leyes mecánicas. Sin embargo, también somos seres racionales que pertenecemos al mundo inteligible y, por lo tanto, podemos resistir las fuerzas mecánicas e ir en una dirección diferente a la que impone la necesidad natural. Para recurrir de nueva a la imagen de Aristóteles, "deberemos tirar de nosotros mismos en sentido contrario […] como hacen los que quieren enderezar las vigas torcidas" (EN 1109b 3-5), es decir, aunque nuestra naturaleza vaya en una dirección, gracias a nuestra condición racional podemos resistir este impulso mecánico y actuar según la ley moral. Es por eso que actuar por deber es el rasgo propio de un agente que habita en un mundo de causas.

Con este análisis se concluye que, a la metafísica mecanicista, Kant superpone la idea de deber. Dado que la concepción metafísica del mundo moderno carece de valor, Kant lo introduce a través de la idea de deber y de la obligación. Este es el sentido en que el deber se im-pone sobre la naturaleza. En esto consiste la solución kantiana a la imagen mecanicista del mundo. Lo interesante es que para hacer compatible esta imagen del mundo con la idea kantiana de deber es necesario comprometerse, primero, con la distinción entre la naturaleza y el deber, y segundo, si el deber ha de tener la carga normativa que Kant le otorga, entonces también surge un compromiso con el requisito prescriptivista: la naturaleza (mecánica) es como es y esto no resta en nada el valor (moral) que tiene el deber.

Existe entonces una relación conceptual muy estrecha entre la metafísica del mundo moderno y la distinción kantiana entre el deber y la naturaleza. No busco afirmar que la distinción misma solo pueda ser consistente con la visión moderna del mundo, ni tampoco afirmo que esta visión implique la distinción deber/naturaleza. No creo que esto sea así, pero no argumentaré en esa dirección. Mi conclusión es más limitada y está enmarcada dentro de la filosofía kantiana: la distinción deber/naturaleza, que es el presupuesto del requisito prescriptivista kantiano, está soportada por la concepción moderna del mundo. En efecto, dentro del paradigma kantiano, es en este mundo mecanicista y desencantado donde tiene sentido la distinción entre cómo somos los seres humanos y cómo debemos ser.

Mi interés en este análisis metafísico se explica por sus implicaciones prácticas y morales. Considero que, desde la perspectiva kantiana, a esta imagen moderna del mundo le corresponde una concepción particular de la condición del ser humano –una antropología–, así como una determinada concepción de la condición de ser-agente. Asumo, entonces, que existe una dependencia teórica entre una visión metafísica y sus consecuencias antropológicas y agenciales. Pienso que la forma en la que el mundo *es* –o mejor, la forma en la que concebimos que es– configura nuestra imagen de la naturaleza humana, necesariamente nuestras concepciones metafísicas afectarán en alguna medida nuestras concepciones sobre la condición humana. Adicionalmente, al contar con una concepción antropológica, también es posible derivar una concepción de la agencia consistente con ella, pues cómo sea definida nuestra condición humana determina, al mismo tiempo, la concepción de la naturaleza de nuestras acciones: la forma de comprender lo que hacemos y lo que debemos hacer depende de la concepción que tengamos sobre el tipo de seres que somos; si fuéramos otro tipo de seres diferentes, seguramente actuaríamos de otras maneras que no tendrían sentido dentro de nuestra concepción de la condición humana[38].

[38] Un ejemplo de esto lo ofrece el mismo Kant con la figura del santo: si fuéramos santos, la naturaleza de nuestras acciones morales no estaría configurada por el deber, sino por el querer (*F* 4:440). Esto ilustra que la naturaleza de nuestra voluntad determina la naturaleza de nuestras acciones.

3.3 El perspectivismo kantiano

¿Cuál es la concepción antropológica y de la agencia que se deriva de la imagen desencantada del mundo? Aun cuando la concepción antropológica esté estrechamente relacionada con la concepción de la agencia, se trata de dos ideas diferentes y, por tanto, las trataré separadamente.

¿Qué concepción del ser humano se deriva de la imagen del mundo moderno? Inmediatamente habría que decir que si el mundo es solo un conjunto de eventos mecánicos, resulta que el ser humano se reduciría a un objeto más en esta cadena causal, un objeto sometido a las leyes naturales. Resulta entonces que esta imagen desencantada del mundo, por sí misma, y al menos para Kant, excluiría o eliminaría la posibilidad de actuar independientemente de la cadena causal. En otras palabras, si solo atendemos al mundo como un mecanismo, el ser humano quedaría reducido a un objeto. Esto es problemático, al menos para Kant, pues la idea misma de que somos objetos no le da sentido a nuestra experiencia humana. Si los seres humanos fuéramos solo objetos, nociones como la agencia, la libertad, la responsabilidad, el libre albedrío, entre otras, no tendrían sentido. No sobra decir que estas nociones, para Kant, también son absolutamente ineludibles e irrenunciables[39].

[39] Kant consideraba que los seres humanos no podemos desprendernos de una noción como la de la libertad (y consecuentemente de las nociones relacionadas: responsabilidad y evaluación moral). Se me ocurren al menos dos razones para justificar esta idea. La primera es que la naturaleza de nuestra razón humana no puede eludir las preguntas por lo incondicionado: "la razón humana tiene, en un género de sus conocimientos, el singular destino de verse agobiada por preguntas que no puede eludir, pues le son planteadas por la naturaleza de la razón misma, y que empero tampoco puede responder; pues sobrepasan toda facultad de la razón humana" (*C1* AVI). La segunda razón es que, tal y como se concibe el proyecto kantiano de fundamentar la moral, nuestras prácticas morales humanas tienen como condición de posibilidad la idea de libertad. Sin este presupuesto no podríamos darle sentido al *hecho* de que nos reconocemos como agentes morales: "hemos de presuponerla [la libertad], si queremos imaginarnos un ser en cuanto ser racional […] como dotado de una voluntad" (*F* 4:449).

¿Cómo resolver esta dificultad sin negar la metafísica mecanicista? Kant intenta darle un lugar especial al ser humano en este mundo de causas. Su solución consiste en suponer "la existencia de algún tipo de fuerza motivacional que es externa al carácter específico del agente y que tenga, sin embargo, el poder para eventualmente vencer sus deseos e intereses. […] La *'voluntad'* es el nombre que la tradición kantiana ha reservado para este tipo de fuerza motivacional" (Corbí, 2012, p. 41; énfasis en el original). En otras palabras, la solución consiste en suponer una fuerza de la agencia que sea externa al mundo mecánico, pero que, aún así, pueda operar causalmente[40]. De esta manera, el ser humano a la vez que es un objeto sometido a las leyes de la naturaleza, es también un agente en un mundo de causas. Mientras la naturaleza humana hace parte del mundo mecánico, la voluntad está por fuera de la causalidad mecánica y puede afectar o llegar a determinar a la naturaleza humana.

La forma en la que esta voluntad procede no es a través de una determinación mecánica. La voluntad –la razón práctica (*F* 4:441; 4:448)– no pertenece a este mundo de fuerzas. La forma de la determinación que procede de la voluntad es normativa, es decir, se manifiesta a través del deber y de la obligación. El deber se im-pone sobre la naturaleza humana, aunque esta sí está sometida a los influjos causales.

Kant entonces busca hacer compatible la metafísica del mundo moderno con la idea de que el ser humano no está sometido a una forma de causalidad mecánica. Sin embargo, las consecuencias filosóficas de esta solución son muy radicales. Como decía, Kant no puede negar que los seres humanos somos objetos de este mundo mecánico; pero para no reducir al ser humano a la condición de objeto, él busca sostener que, en algún sentido, los seres humanos no dependemos (enteramente) de esta determinación mecánica; o en otras palabras, somos independientes con

[40] Cabe anotar que esta fuerza motivación que se le atribuye a la voluntad es muy sospechosa porque se trata de una "potencia no-mecánica que puede interactuar con las fuerzas mecánicas del agente" (Corbí, 2012, p. 41). En efecto, la naturaleza humana, por un lado, y de la voluntad humana, por otro, crean una brecha que demanda una explicación sobre cuál pueda ser su relación.

respecto a las leyes mecánicas. Esta idea lo lleva a proponer algo que ya había mencionado antes: el *punto de vista [Standpunkt]*. Esta idea consiste en considerar al ser humano desde dos puntos de vista o perspectivas distintas, como *noúmeno* y también como *fenómeno* (*F* 4:450; 4:452; 4:455; 4:458).

> El hombre, como ser *natural* dotado de razón (*homo phaenomenon*), puede ser determinado por su razón, como *causa*, a realizar acciones en el mundo sensible, y con esto todavía no entra en consideración el concepto de obligación. Pero él mismo, pensado desde la perspectiva de su *personalidad*, es decir, como un ser dotado de *libertad* interna (*homo noumenon*), se considera como un ser capaz de obligación y, particularmente, de obligación hacia sí mismo (la humanidad en su persona). (MC 418; énfasis en el original)

Lo anterior enuncia la tesis del perspectivismo kantiano*:* desde la perspectiva de la naturaleza somos seres sensibles sometidos a la causalidad del mundo; pero desde la perspectiva de la razón, somos seres capaces de respetar la ley moral y actuar por deber [*aus Pflicht*].

Con esta consideración es posible perfilar la concepción antropológica kantiana que se deriva de la metafísica moderna. El aspecto fenoménico del ser humano se corresponde con su naturaleza corporal, una naturaleza que está sometida a la causalidad y a las leyes naturales. El aspecto nouménico se corresponde con la razón, y ella es la fuente de la obligación y el deber. La consecuencia de la imagen del mundo moderno es que el ser humano termina dividido: es condición natural y es condición racional.

La noción de deber es tan heterogénea a la naturaleza sensible del ser humano que Kant termina por escindir la condición humana para hacerle un espacio a las exigencias del deber en el mundo de causas. La conclusión que se sigue de esto es que Kant está comprometido con una antropología filosófica que se define a partir de la *división de la condición humana.*

Con este análisis quiero llevar al lector a ver de qué modo, en el marco del paradigma metafísico

moderno, el requisito prescriptivista presupuesto por el rigorismo está comprometido con la tesis de que el ser humano es un ser nouménico y también un ser fenoménico. En otras palabras, y de modo más esquemático: (i) el rigorismo moral solo tiene sentido si supone la distinción entre deber-ser y naturaleza humana. (ii) Esta distinción, a su vez, está soportada sobre la metafísica del mundo moderno. (iii) Y a patir de esta imagen del mundo se deriva una concepción del ser humano dividido en dos condiciones: la nouménica y la fenoménica.

3.4 La concepción del yo dividido

Kant acepta la brecha entre deber y naturaleza, y acepta también la división de la condición humana arriba descrita. Tal y como interpreto su posición, esta división se reproduce varias veces a través de otras de sus distinciones conceptuales: razón y sensibilidad; moralidad y felicidad; deber e inclinación; buena-voluntad y voluntad-humana. Todas estas parejas conceptuales dependen de la división de la condición humana.

Una perspectiva antropológica en la que la condición humana está dividida, evidentemente, dará lugar a una concepción particular de la agencia. Retomo la idea de que nuestra condición como seres humanos determina la naturaleza de nuestras acciones: lo que hacemos, cómo lo hacemos y hasta cómo debemos hacerlo depende sustancialmente del tipo de seres que creemos ser. En esta última sección el análisis estará enfocado en caracterizar el modelo de agencia que se deriva de la condición dividida del ser humano.

¿Cómo derivar una noción de la agencia a partir de la distinción antropológica entre noúmeno y fenómeno? Paso primero a explicar la idea general y luego intentaré desarrollarla. La idea general es la siguiente: cuando atendemos exclusivamente al punto de vista fenoménico estamos sometidos a las leyes naturales. En este sentido, nuestras acciones están determinadas por leyes que no están bajo nuestro control. Cuando entendemos al ser humano como parte de este entramado causal, esto es, en tanto que fenómeno, entonces lo vemos reducido a la condición de objeto. Bajo esta

interpretación, la condición fenoménica del ser humano elimina la posibilidad de agencia. En constraste con esto, cuando atendemos a la condición nouménica, el panorama cambia. Como seres nouménicos, los seres humanos se independizan de la causalidad de las leyes naturales. Y al librarse de este condicionamiento causal sus acciones están en alguna medida bajo su control. Esta es otra forma de decir que los seres humanos, en tanto seres nouménicos, somos libres. Ahora bien, al ganar libertad y (algún) control sobre sus acciones, evidentemente el ser humano adquiere (cierta) condición de agencia. Por esta razón, la agencia se valida únicamente desde la perspectiva nouménica.

Esta presentación general es útil para explicar (i) que la distinción entre noúmeno y fenómeno está conceptualmente asociada al concepto de libertad, y (ii) que la noción de libertad, a su vez, está conceptualmente asociada a la noción de agencia. Es por esta razón que, según mi interpretación, existe una relación entre la distinción noúmeno/fenómeno y la noción de ser agente. Sin embargo, como veremos, los detalles del argumento son más complicados porque Kant ofrece diferentes definiciones del concepto de libertad. En todo caso, el punto importante no cambia: a partir del concepto kantiano de libertad, que se sustenta en la distinción entre noúmeno y fenómeno, es posible derivar una noción de la agencia. Con esta idea en mente, expondré el argumento con más detalle.

3.4.1 El concepto kantiano de libertad en la *Fundamentación de la metafísica de las costumbres*

La idea de libertad es uno de los ejes centrales de la filosofía práctica kantiana. Además se trata de un concepto que por sí mismo daría para una investigación independiente y muy exigente. Por lo pronto, advierto que el sentido de este concepto varía a lo largo de la obra de Kantt. Para los objetivos de mi argumento, solo presentaré los aspectos generales del concepto, teniendo en cuenta sus variaciones. Me centraré primero en la definición dada en la *Fundamentación de la metafísica de las*

costumbres y más adelante examinaré la variación en la *Metafísica de las costumbres*.

En la *Fundamentación* Kant sostiene que la causalidad opera no solo en el mundo de los fenómenos, sino también en el mundo nouménico. Sin embargo, se trata de tipos distintos de causalidad. Mientras en el mundo fenoménico opera la causalidad mecánica, en el mundo nouménico, al que pertenece la voluntad, opera un "tipo de causalidad de los seres racionales" (*F* 4:446). La voluntad es justamente este tipo de causalidad que no opera según las leyes mecánicas naturales. Ella tiene una propiedad especial: la libertad. Y esta propiedad permite que la voluntad racional "sea eficiente independientemente de causas ajenas que la *determinen*" (*ibid.;* énfasis en el original).

La libertad, por definición, contrasta con la necesidad natural, la cual es la propiedad de la causalidad de los seres irracionales. Estos "ven determinada su actividad por el influjo de causas ajenas" (*F* 4:446). En este sentido, los animales no humanos, al igual que los objetos, están sometidos a una causalidad sometida a la necesidad natural, su actividad no depende de ellos, sino de las leyes naturales. Como se ve, los seres racionales tenemos la propiedad de ser libres, mientras que los seres irracionales no pueden llegar a serlo.

Según Kant, esta "independencia de causas ajenas" es una definición negativa de libertad, porque se refiere a la condición de no estar determinado. Sin embargo, la definición no agota todo el concepto de libertad. Kant propone también una definición positiva de libertad. Podría decirse que las dos definiciones son dos sentidos complementarios del mismo concepto. La libertad negativa hace referencia a la independencia del agente con respecto al mundo mecánico, es decir, el agente es *libre de* determinaciones ajenas. La libertad positiva, en cambio, se refiere a que la voluntad del agente es *libre para* "ser una una ley para sí misma" (*F* 4:447).

En su sentido positivo, la libertad también está sometida a leyes. Es por esto que Kant afirma que la libertad no es "anómica" (*F* 4:446), pues una voluntad libre sin leyes sería un absurdo, una voluntad sin principios sería igual a ninguna voluntad en absoluto (ver §5.1). Sin embargo, la

voluntad opera según "leyes inmutables" que no provienen del exterior de la voluntad. Todo lo contrario, estas leyes provienen de la voluntad misma, son leyes que la voluntad se da a sí misma. En pocas palabras, lo que se describe aquí es el concepto kantiano de autonomía: "la propiedad de la voluntad de ser una ley para sí misma" (*F* 4:447). La voluntad libre es entonces una voluntad autónoma[41]. Este es el sentido positivo del concepto de libertad que propone Kant.

Cabe aclarar que a la voluntad humana no le es intrínseca la propiedad de la libertad, y por tanto, la voluntad no siempre es autónoma. Puede suceder que la voluntad actúe movida por leyes que estén determinadas en "algún otro lugar" (*F* 4:441), leyes generadas por móviles externos a la voluntad misma. Cuando la ley viene dada por el objeto externo de la voluntad, entonces la voluntad es heterónoma. La voluntad es heterónomca, por ejemplo, cuando actúa motivada por el deseo de un objeto externo, porque, en un caso como este, la máxima que la determina es ajena a ella misma. "La voluntad no se da a sí misma la ley, sino que ésta le viene dada por un impulso ajeno mediante una naturaleza del sujeto afinada para la receptividad de dicho impulso" (*F* 4:444).

3.4.2 La noción de libertad en la *Metafísica de las Costumbres*: ¿cómo imputar acciones no morales?

La definición de libertad positiva, tal y como se expresa en la *Fundamentación*, genera un serio problema. El problema surge al separar de un modo tan tajante la condición humana entre noúmeno y fenómeno. Si estos dos puntos de vista son tan heterogéneos y responden a causalidades

[41] El argumento de Kant va más allá de este punto. Según el autor, que "la voluntad sea en todas las acciones una ley para sí misma" es básicamente el principio de "obrar conforme a una máxima que la voluntad pueda tener como ley universal", ya que la voluntad asentiría universalmente a una máxima que ella misma origina. Dicho esto, Kant concluye, entonces, que la autonomía de la voluntad, esto es, una voluntad libre, es exactamente lo mismo que una voluntad regida por la ley moral (*F* 4:446). Este argumento muestra entonces que la libertad y la moralidad se identifican: un agente libre (autónomo) actuaría moralmente. Sin embargo, esta conclusión es muy problemática, especialmente porque en las ocasiones en las que no actuamos moralmente no seríamos libres, y por tanto no podríamos ser responsables.

tan distintas, parece que se pierde todo tipo de conexión entre ellas; es como si en vez de una división de la condición humana, se tratara de dos seres distintos: uno racional-autónomo y otro corporal-heterónomo. Una versión filosófica del Extraño caso del Dr. Jekyll y Sr. Hyde. Bajo esta interpretación, habría que aceptar que las acciones causadas por deseos o por inclinaciones no podrían ser atribuidas al agente-racional. En efecto, como seres fenoménicos, estamos sometidos a las leyes de la naturaleza. Esto significa que nuestra voluntad puede estar determinada por objetos del mundo exterior que causen en nosotros un deseo o una inclinación hacia ellos. En esa medida, nuestra voluntad estaría condicionada externamente y, por tanto, resultaría siendo dependiente del mundo, resultaría siendo una voluntad heterónoma. En una situación semejante se pierde la independencia causal y, por tanto, el agente no puede autodeterminar su voluntad, porque está sometido a los influjos causales. Dada esta situación, el agente no puede responder por acciones que hayan sido causadas por móviles del mundo fenoménico. A esto se le conoce como el *problema de la imputabilidad de las acciones no morales* (Caimi, et al., 2016, p. 46).

Kant es consciente de este problema. Su solución es presentada en la *Metafísica de las costumbres* y consiste en mostrar que la condición nouménica influye, o puede influir, en la condición fenoménica. De esta manera, sería posible que el agente libre pudiera tener algún tipo de control o influencia sobre lo que acontece en el mundo fenoménico y, por tanto, las acciones que podrían llamarse heterónomas bien pueden imputarse al agente.

Si Kant logra mostrar que la condición nouménica influye en la condición fenoménica, estaría demostrando, en otras palabras, que los agentes actuamos en un mundo de causas. Para probar esta tesis, sin embargo, necesita mostrar (i) que de alguna manera la agencia supone independencia del mundo mecánico, y (ii) también necesita mostrar cómo actuar libremente en el mundo de causas. Si Kant logra darle sentido a la libertad del agente en el contexto de la metafísica moderna, entonces puede darle sentido a la idea de que podemos ser autores de nuestras propias acciones.

En la *Metafísica de las costumbres* Kant distingue entre la voluntad [*Wille*] y el arbitrio [*Willkür*]. A la

voluntad le corresponde generar leyes morales; por su parte, el arbitrio genera las máximas. Kant define una máxima como el principio subjetivo del querer (*F* 4:401n) que solo tiene validez para un sujeto individual. Las máximas tienen, entonces, la característica de no ser objetivamente válidas. Por el contrario, la ley moral es un principio objetivo del querer que es, por tanto, válido para toda voluntad racional (*ibid.*). Mientras la ley moral es tal que la voluntad de un ser racional se la impone a sí misma como ley universal (Caimi, et al., 2016, p. 284), la máxima es simplemente el principio de acción que un agente elige para guiar su acción, dada una determinada situación.

A partir de esta distinción, Kant sostiene que la voluntad [*Wille*] no puede ser libre ni no-libre pues esta facultad solo se refiere a la legislación que han de seguir las máximas de acción. Y por esta razón, la voluntad misma no es susceptible de ninguna coerción y ningún influjo sensible puede afectarla; ella no tiene ninguna relación con el aspecto fenoménico del ser humano. Es por esto que la voluntad ha sido descrita como la "razón práctica misma" (*MC* 226). El arbitrio [*Willkür*], en cambio, hace alusión a las acciones concretas y, por lo tanto, sí guarda una relación con la parte sensible del ser humano. Para Kant, por tanto, solo podemos denominar libre al arbitrio (*ibid.*).

Cuando Kant está hablando de la voluntad (*Wille*) y de su capacidad legisladora, se está comprometiendo con la idea de libertad como autodeterminación. Kant afirma que las leyes morales proceden de la voluntad (*MC* 226). Esto significa que la voluntad solo asiente universalmente a las máximas que ella misma origina, pues no podría asentir universalmente a máximas externas a ella, ya que hacerlo no tendría validez para una voluntad pura. En este sentido, la voluntad tiene la propiedad de ser una ley para sí misma, es decir, ella puede autolegislarse (esta es exactamente la noción de autonomía —libertad positiva— propuesta en la *Fundamentación*). En contraste, cuando se introduce la idea de libre arbitrio (*Willkür*), Kant en realidad está hablando de libertad como *posibilidades alternativas*: como la capacidad que tiene el arbitrio para elegir entre máximas de acción morales y no-morales. La voluntad [*Wille*] no puede elegir entre distintas posibilidades, pues, como se ha dicho, su legislación moral es necesaria. En cambio, el arbitrio tiene la posibilidad de actuar no-moralmente.

La distinción entre voluntad y arbitrio genera algunos problemas conceptuales que aquí apenas menciono. En primer lugar, Kant encuentra que la noción de libre arbitrio es una mezcla entre un concepto nouménico y una experiencia fenoménica. Esto, por supuesto, no es consistente con el perspectivismo kantiano[42]. Adicionalmente, afirmar que la voluntad no es libre ni no-libre, o afirmar que un ser humano puede elegir máximas no morales parece que contradice tesis fundamentales de la *Fundamentación*, en donde la voluntad humana aparece como libre cuando es autónoma, y la libertad se identifica con la sujeción a ley moral (Caimi, et al., 2016, p. 46). Dejando esto de lado, lo interesante de la distinción entre voluntad [*Wille*] y arbitrio [*Willkür*] es que la voluntad siempre legislará de acuerdo con la ley moral, nunca se alejará de los dictados morales. En cambio, el libre arbitrio, al menos en nuestra experiencia fenoménica, nos permite elegir máximas morales y también aquellas no morales. Esto significa que, después de todo, sí podemos elegir, en tanto agentes, acciones no morales.

Korsgaard ofrece una interpretación interesante de la forma en la que opera el libre arbitrio: se puede decir que el agente hace uso de su libre arbitrio cuando reflexivamente "la mente asiente al deseo antes de poder actuar conforme a él" (Korsgaard, 2000a, p. 122). Esto significa que el agente puede tener una máxima de acción que satisfaga el deseo y puede también asentir libremente a ella en virtud de su libre arbitrio. "Entonces, aunque podemos hacer lo que el deseo exige, lo hacemos libremente" (*ibid.*). Obviamente, si adoptamos el deseo como nuestra máxima de acción, no estamos en el terreno moral, ya que un deseo no constituye por sí solo una máxima que se atenga al imperativo categórico. Sin embargo, aunque actuar de acuerdo con el deseo no se

[42] Kant reconoce que no es posible definir teóricamente el libre arbitrio como la facultad para elegir entre máximas morales y no morales (*MC* 226). Es cierto que, desde la perspectiva fenoménica, el libre arbitrio nos da ejemplos de esto y constatamos que es posible elegir máximas de acción contrarias a la moral. Pero desde la perspectiva nouménica, no es posible que "los fenómenos hagan comprensible un objeto suprasensible" (*ibid.*), esto es, no es posible definir el libre arbitrio en virtud de su relación con la sensibilidad. De manera que el libre arbitrio parece solo una experiencia fenoménica que se contrapone a la voluntad nouménica, ya que esta última sí actúa independientemente de cualquier fenómeno causal.

corresponda necesariamente con un acto moral, sí que puede ser un acto libre, pues la voluntad no se ha dejado llevar por el deseo, sino que ha asentido a una máxima que busca satisfacerlo. Para usar expresiones comunes, lo que se describe acá es la diferencia entre "me ganó el deseo" y "elegí este deseo". Aunque en ambos casos el agente satisface el deseo, solo en el segundo hubo una máxima de acción que fue producto de la reflexión. Lo que importa aquí es que el agente puede ejercer su libertad —o para ser más precisos, su arbitrio— sobre el terreno fenoménico[43]. De esta manera, la escisión entre noúmeno y fenómeno no es insalvable: en la medida en que es posible elegir, incluso en el terreno fenoménico, podemos darle sentido a la idea de ser-agentes en un mundo de causas[44].

A partir de esta interpretación resulta que la libertad del arbitrio del agente no tiene una función

---

[43] Existe una objeción a esta interpretación. Si las acciones humanas están determinadas causalmente en este mundo mecánico, entonces la supuesta libertad de elegir la máxima de acción es tan solo una ilusión (Korsgaard, 2000a, p. 122). Después de todo, si fuera posible predecir nuestra acción porque contamos con la información causal suficiente, entonces no tendría sentido elegir la máxima de acción ya que de todas formas terminaremos haciendo lo que la cadena causal determine.
Según Korsgaard, esta preocupación puede reformularse de la siguiente manera: "si tuviéramos la seguridad de estar determinados o supiéramos cómo estamos determinados, no podríamos actuar, no actuaríamos, o bien actuaríamos de manera diferente" (*ibid.* p. 123). Pero a decir verdad, el hecho de que eventualmente podamos predecir nuestra conducta según leyes naturales, no significa que seamos menos libres, pues saber cuál será mi conducta no elimina el hecho de que yo elija actuar según la máxima de acción.
Desde la perspectiva externa, desde la del científico con una concepción del mundo moderno, mi conducta podría ser perfectamente predecible porque bastaría con identificar todas las cadenas causales que me llevarán a actuar de tal o cual forma. Y con todo, según la interpretación que hace Korsgaard de Kant, eso no invalida ni suprime que el agente, a través de su libre arbitrio, elija libremente que su máxima de acción sea aquella que está determinada por la cadena causal, o bien elija que su máxima de acción sea contraria a esta cadena causal.
[44] El concepto de libertad, al menos bajo esta interpretación de Korsgaard, está planteado en términos prácticos: el conocimiento de las cadenas causales puede afectar o influir sobre la elección, pero no la suprime. La libertad que se descubre en la reflexión no es una propiedad teórica que también puedan ver los científicos que estudien las deliberaciones del agente desde fuera, desde su perspectiva de tercera persona. Es desde dentro, desde la perspectiva deliberativa, desde donde notamos que nuestros deseos nos ofrecen sugerencias que podemos seguir o ignorar (Korsgaard, 2000a, p. 124).

explicativa-científica. La libertad no es una propiedad del mundo, ni hace parte de la explicación causal de las acciones de un agente. Por el contrario, la libertad del arbitrio sirve para describir la condición en la que nos encontramos cuando reflexionamos sobre qué hacer; y en este sentido, esta libertad posibilita nuestra capacidad de tener razones para actuar: elegir libremente un deseo como máxima de acción significa que el deseo es la razón que tengo para actuar; y esto es independiente de que el el objeto del mundo haya causado este deseo en mi facultad desiderativa.

Por esta razón, dice Korsgaard, no existe una incompatibilidad entre el determinismo y la libertad del arbitrio (2000a, p. 123). Y esto, a su vez, se traduce en que no existe una inconmensurabilidad entre la condición fenoménica y la condición nouménica del ser humano, pues aunque estemos condicionados causalmente, podemos siempre elegir si este condicionamiento será la máxima que dirija nuestra acción.

Esta explicación resuelve el problema de la imputabilidad de las acciones no morales, pues si el ser humano es libre incluso para elegir máximas no morales o máximas que buscan satisfacer deseos o inclinaciones, evidentemente podemos atribuirle responsabilidad por las acciones derivadas de estas máximas. En otras palabras, el agente kantiano tiene (cierto) control incluso sobre su condición fenoménica.

3.4.3 La concepción del yo dividido

Soy consciente de las contradicciones conceptuales que se presentan entre la *Fundamentación* y la *Metafísica de las costumbres* con respecto al concepto de la libertad. Por suerte no necesito resolver esta tensión, ya que el objetivo de este capítulo no es explicar la teoría kantiana de la libertad. Lo que me interesa es la pregunta por la noción de agencia que se desprende de estas distintas formas de pensar la libertad.

Considero que al tener distintos conceptos de libertad evidentemente surgirán distintas nociones de agencia. La relación entre libertad y agencia estriba en que el alcance de las acciones (lo que alguien hace) depende del grado de libertad que tenga (lo que puede hacer). La pregunta, entonces, es ¿qué nociones de agencia se desprenden de estas definiciones de libertad?

La respuesta que he dado a esta pregunta se articula en dos momentos. En un primer momento, a partir de la *Fundamentación* se concluye que la distinción noúmeno/fenómeno está aparejada con la distinción entre libertad y necesidad natural, respectivamente. En tanto que ser nouménico, la voluntad en el ser humano es libre (en sentido negativo y positivo). De esta manera, el ser humano se independiza de su condición fenoménica y su voluntad es autónoma. Sin embargo, esta interpretación nos deja, como vimos, con el problema de la imputabilidad de las acciones no morales. En un segundo momento, y para resolver este problema, apelo a la *Metafísica de las costumbres*. La solución kantiana consiste en introducir una diferencia entre voluntad (*Wille*) y libre arbitrio (*Willkür*). Según esto, los seres humanos podemos elegir incluso máximas que no son morales. Esto permite que, en el mundo fenoménico, haya libertad de arbitrio y, por tanto, agencia. Con este argumento se explica cómo es posible ser un agente en un mundo mecánico.

Con los elementos que he mostrado hasta aquí se puede derivar un modelo kantiano de la agencia. En mi interpretación, la concepción antropológica kantiana, según la cual la condición humana está dividida entre lo nouménico y lo fenoménico, se traduce en un modelo de la agencia. Este modelo lo extraigo justamente de la relación que se puede dar entre la condición nouménica y la fenoménica del ser humano. Por supuesto, la relación depende del tipo de interpretación que adoptemos, ya sea la derivada de la *Fundamentación* o la que se deriva de la *Metafísica de las costumbres.* En cualquier caso, la agencia tiene su lugar exclusivamente en la condición nouménica del ser humano, la diferencia depende de la relación que tenga la agencia-racional con la condición fenoménica.

En la interpretación de la *Fundamentación,* la libertad positiva se entiende como la propiedad de la

voluntad que consiste en la capacidad para que ella se de las leyes morales a sí misma, es decir, la voluntad libre es equivalente a una voluntad que se autodetermina. Esta es la caracterización de la autonomía kantiana. Sin embargo, esta propuesta implica que la única forma de ser agente es cuando la voluntad es autónoma, esto es, cuando la voluntad establece principios de acción que se ajustan al modelo kantiano de deliberación racional. En cambio, en la condición de heteronomía, cuando la voluntad está determinada por causas externas a la voluntad misma, no hay agencia posible. Bajo esta interpretación, la agencia excluye una parte de la condición de ser humano, excluye su condición fenoménica.

> En cuanto fuentes de necesidades, las inclinaciones mismas distan tanto de albergar un valor absoluto para desearlas por ellas mismas, que más bien ha de suponer el deseo universal de cualquier ser racional el estar totalmente libre de ellas. (F 4:428)

En la interpretación de la *Metafísica de las costumbres,* gracias al concepto de libre arbitrio [*Willkür*], el agente puede tener algún tipo de control sobre esa parte fenoménica. Evidentemente, el privilegio de la razón prevalece. Ser agente ahora incluye controlar esa parte no racional del ser humano. La formulación de Korsgaard es bastante elocuente: "cuando usted delibera, es como si hubiera algo por encima de todos sus deseos, algo que es *usted,* y que *elige* conforme a cuál deseo actuar. Esto significa que considera que el principio o ley por el cual determina sus acciones lo expresa a *usted mismo*" (Korsgaard, 1996, p. 129; énfasis en el original).

Lo que busco extraer de aquí es que la agencia kantiana o bien excluye la parte fenoménica del ser humano, o bien ejerce un control sobre ella. En ambos casos, Kant está comprometiéndose con la idea de que nos encontramos divididos: "uno debe ser dos" (Korsgaard, 1996, p. 134n). Y esto, en pocas palabras, es lo que Josep Corbí (2012) denomina la *Concepción del Yo Dividido* [CYD]. La concepción del yo dividido es el nombre para describir el modelo agencial kantiano según el cual se supone que una parte del agente es pasiva y otra activa (Corbí, 2012, p. 41). La parte pasiva está compuesta por un conjunto de fuerzas naturales mecánicas que no responden a ningún tipo

de normatividad, y que además tienden a distorsionar nuestras habilidades deliberativas-racionales (*ibid.*). Aun cuando esta parte del yo tenga una fuerza y una energía —al menos en sentido mecánico— es considerada una parte pasiva porque está sometida a las determinaciones causales de la naturaleza. La naturaleza corporal del ser humano es la parte pasiva. Por otro lado, la parte activa del agente kantiano es la encargada de deliberar y tomar decisiones a las que, idealmente, la parte pasiva tendría que adecuarse gracias al "esfuerzo de la voluntad" (*ibid.*). El deber es el resultado de la deliberación moral racional y supone el ejercicio activo de una voluntad cuya función es determinar y guiar la parte pasiva.

Un rasgo de la concepción del yo dividido es que introduce la idea de que los seres humanos estamos divididos en "partes". El aspecto fenoménico del que habla Kant corresponde a la parte pasiva, y el aspecto nouménico se refiere a la parte activa. Mientras que Kant caracteriza la división de la condición humana en términos de aspectos o perspectivas, la concepción del yo dividido se compromete con que los seres humanos estamos estructuralmente divididos. Hablar de partes del yo es una tesis más fuerte que hablar de aspectos. La ventaja interpretativa de la concepción del yo dividido es que se adecúa mejor a la metafísica del mundo moderno porque a la vez que le da contenido al aspecto fenoménico (la parte pasiva obedece a leyes naturales), también enfatiza la idea de que el aspecto nouménico no hace parte del mundo mecánico (la parte activa es distinta/externa a las leyes naturales).

La concepción del yo dividido, además, supone un cierto ejercicio de control (*Control Constraint*) (Corbí, 2012, p. 41). Se trata de otra forma de insistir en la idea de que para lograr la relación moralmente adecuada entre la naturaleza y el deber es necesario que la parte racional controle la parte pasiva. Solo de esta manera el ser humano puede llegar a ser un agente (incondicionado) que se superpone a a su naturaleza condicionada.

Insisto en que la concepción del yo dividido es un modelo de la agencia, y como tal, tiene implicaciones en la comprensión de la acción humana. Resulta que, a través del ejercicio de

control, este modelo introduce un tipo de jerarquía: la parte activa, en tanto ha de ejercer control, es superior a la parte pasiva. De manera que la parte activa del ser humano, su parte racional, posee ciertos privilegios (morales, metafísicos) con respecto a la parte fenoménico. Esta característica evidencia que la CYD es una instancia de las denominadas teorías del yo real*:*

> De acuerdo con las teorías del yo real, uno podría pensar en las acciones como derivadas de dos partes distintas del agente. Un parte, el Yo Real, pertenece más profundamente al agente que la otra, y entonces, tiene un estatus privilegiado en la vida metafísica y moral del agente […]. La otra parte del yo, que podría ser llamada el Falso Yo, no tiene este estatus especial. (Arpaly, & Schroeder, 1999, p. 165)

Es verdad que la palabra *real* hace ruido, ¿acaso la parte pasiva es irreal? ¿Qué significa que el yo sea real? Para precisar, no quiero decir que el yo fenoménico sea irreal o que sea una ficción; el adjetivo *real*, en el contexto de la filosofía kantiana, se refiere a que el yo real se corresponde al yo racional o "yo auténtico" (*F* 4:458). Este es el referente de la agencia moral, mientras que el yo fenoménico —que posee tanta realidad como el yo racional— está excluido de la moralidad, excepto como algo frente a lo cual hay que oponer resistencia.[45] El yo real, en la filosofía moral kantiana, es entonces el yo auténtico o yo racional. Es este el que ejerce o debe ejercer control normativo sobre la parte fenoménica del ser humano.

Lo que me interesa de la teoría del yo real o auténtico es que enfatiza el hecho de que la concepción

[45] "Dentro de dicho mundo [inteligible] la ley es dada exclusivamente por la razón, por una razón ciertamente pura e independiente de la sensibilidad, sabiendo igualmente que, como en ese mundo él solo es un auténtico yo en cuanto inteligencia (en cambio como hombre sólo es fenómeno de sí mismo), le corresponde inmediata y categóricamente, de suerte que aquello hacia lo cual le incitan inclinaciones e impulsos (toda la naturaleza del mundo sensible) no puede causar quebranto alguno a las leyes de su querer en cuanto inteligencia, hasta el punto de que no se responsabiliza de esas inclinaciones e impulsos y no los imputa a su *auténtico yo*, esto es, a *su voluntad*, aunque sí se responsabilice de la indulgencia que pueda prodigar hacia ellas, cuando les otorga un influjo sobre sus máximas en detrimento de la ley racional de la voluntad" (*F* 4:458).

del yo dividido viene aparejada con una tesis normativa: la CYD no solo se refiere a la descripción de la división de la condición humana. Este modelo de la agencia, al ser una instancia de la teoría del yo real, está comprometido con la idea de que una de las partes del ser humano, la parte racional, es la parte más importante en contraste con la parte pasiva, y por lo tanto, opera como criterio o patrón normativo frente al yo fenoménico-pasivo.

La CYD es entonces un modelo de la agencia que caracteriza la condición dividida del ser humano y que supone además una relación normativa entre las partes. Ahora bien, el problema de la CYD no es tanto que describa la condición humana en términos de una división. Si ese fuera el caso, la solución óptima sería un modelo de la agencia en el que el agente fuera una unidad ideal.[46] Y no estoy seguro de que tenga sentido hablar de un agente unificado idealmente (ver §6.1). Más bien, como veremos en el siguiente capítulo, el problema de la CYD es el tipo de relación jerárquica y normativa que se establece entre la parte activa-racional y la parte pasiva-fenoménica. Es en las consecuencias normativas y morales donde se encuentra, en mi opinión, el problema de modelo de la agencia kantiano.

Quisiera terminar haciendo un resumen esquemático del argumento hasta este punto. La conclusión que me interesa defender es que la concepción del yo dividido es el modelo de la agencia que está detrás del rigorismo moral: (i) El rigorismo presupone el requisito prescriptivista. (ii) Este requisito presupone la distinción entre el deber y la naturaleza humana. (iii) Esta distinción está anclada y depende de la imagen desencantada del mundo moderno. A partir de esta imagen de mundo se deriva una concepción antropológica: el ser humano está dividido en su condición, como fenómeno y como noúmeno. Finalmente, (iv) he querido mostrar que esta división entre fenómeno y noúmeno está conceptualmente asociada a la teoría de la libertad. Y el concepto de libertad, a

[46] Agradezco a Alexander Burgos por señalarme este punto en el Seminario de investigación del Doctorado en Filosofía de la Universidad Nacional, en donde presenté una versión previa de este capítulo.

su vez, está asociado a la noción de agencia. Sin embargo, (v) Kant ofrece dos definiciones distintas del concepto de libertad, lo cual implica diferentes nociones de agencia. (vi) Concluyo mostrando que en cualquier caso el modelo agencial kantiano es el mismo: la concepción del yo dividido.

El argumento que presento en este capítulo, insisto, asume que las tesis metafísicas tienen una resonancia en la concepción de lo que es el ser humano y de lo que es ser agente. Cómo entendamos el mundo determina la comprensión de lo que significa ser un ser humano, de cómo nos situamos en él, y de cuál es la naturaleza de nuestras acciones. Pues bien, valiéndome de esta idea, he partido de una tesis moral (rigorismo) y he descendido hasta sus fundamentos metafísicos (imagen desencantada del mundo) para, desde allí, caracterizar una concepción antropológica (división de la condición humana) y una concepción de la agencia (concepción del yo dividido). [47]Con este argumento, creo, puedo concluir que el rigorismo moral kantiano está comprometido con la CYD.

[47] Obviamente este modelo de la agencia no es el *único* compatible con la concepción desencantada del mundo. Por ejemplo, partiendo de que el mundo es solo una relación de causas, podemos pensar que el ser humano en su *totalidad* está sometido a esta causalidad. Por ejemplo, se puede pensar que las emociones y sensaciones son respuestas causales (Hobbes y Hume pueden ser ejemplos de este caso); o que todos los actos humanos están determinados causalmente como lo pensaría un determinista radical. Pero si queremos hacer compatible la distinción entre deber y naturaleza (distinción que está presupuesta por el rigorismo moral) con la imagen desencantada del mundo —tal y como sucede en la filosofía práctica kantiana—, entonces, afirmo, nos comprometemos con la CYD.

4. El problema del rigorismo moral[48]

4.1 Williams: el problema de la integridad personal
4.2 Integridad: carácter y autoindulgencia
4.2.1 Integridad: identidad y carácter
a. Identidad y carácter
b. 'Seguir hacia adelante' y deseos categóricos
c. Proyectos fundamentales y sentido [*meaning*]
d. Amenaza a la integridad: proyectos fundamentales y carácter
e. Muerte práctica
4.2.2 Integridad y autoindulgencia moral
4.3 El problema de la integridad y la concepción del yo dividido

El rigorismo moral representa la consecuencia extrema de la postura kantiana, y es justamente a través de esta lectura rigorista como creo que se pueden entrever mejor los alcances y los límites de esta teoría. En esta línea, en el capítulo anterior intenté mostrar la relación entre el rigorismo y la concepción del yo dividido. Con ello pretendía hacer explícitos los presupuestos metafísicos y el modelo de la agencia que están a la base del rigorismo moral kantiano.

Con esta idea en mente, y una vez descubiertos los presupuestos y compromisos conceptuales y teóricos del rigorismo kantiano, me interesa plantear el problema de la integridad personal. En §4.1, planteo este problema siguiendo la caracterización que hace Bernard Williams. En §4.2, amplío y explico en qué consiste la amenaza a la integridad. Me enfoco para ello en los conceptos de *carácter*, *proyecto fundamental* y *autoindulgencia moral*. En §4.3, explico por qué el rigorismo kantiano constituye una amenaza contra la integridad personal y en qué sentido este no puede darle sentido a la preocupación de una persona por su integridad.

4.1 Williams: el problema de la integridad personal

[48] Una versión previa de este capítulo fue presentada en el *Colloquium - Valencia Philosophy Lab* (octubre 2019) en la Universidad de Valencia. Agradezco a los asistentes de este evento porque sus comentarios y preguntas fueron muy estimulantes y valiosas. Muchas de sus sugerencias e ideas están aquí incluidas.

Si bien creo que la teoría moral kantiana puede ser criticada desde muchos frentes, aquí me interesa centrarme en el problema de la integridad del agente. Este problema fue planteado por Bernard Williams por primera vez en "A Critique of Utilitarianism" (1973a), y posteriormente el mismo autor desarrolló su intuición inicial a lo largo de otros textos, especialmente en "Persons, Character and Morality" (1981d) y en ""Utilitarianism and moral self-indulgence" (1981e). El problema de la integridad consiste en que la moralidad imparcial e impersonal puede llegar a exigirle al agente que actúe de maneras que amenazan su integridad personal. Aunque la crítica de Williams está planteada directamente en contra del utilitarismo, lo que busco en este capítulo es plantear el problema de la integridad en el contexto de la filosofía moral kantiana; adicionalmente, me interesa mostrar de qué manera este problema tiene su raíz en la concepción de yo dividido que está a la base de esta filosofía. Esta formulación, por supuesto, exige varias aclaraciones.

En "A Critique of Utilitarianism", Williams sostiene que una teoría moral como el utilitarismo que obligue a una persona a actuar de acuerdo con principios impersonales e imparciales (p.e. elegir la acción que genere la máxima utilidad posible) falla necesariamente. Su argumento es que una ética imparcial como el utilitarismo —y, como espero mostrar, también como el kantismo— ofrece una imagen incoherente de la vida práctica, pues fuerza una ruptura entre el agente y sus compromisos más sinceros.

Williams utiliza dos ejemplos para ilustrar esta idea. El primero es el caso de un químico que tiene la posibilidad de trabajar en una empresa que investiga para crear armas químicas y biológicas. El químico rechaza vehementemente hacer parte de una investigación para la guerra y, sin embargo, necesita el empleo; su esposa, además, lo alienta a aceptar el trabajo. Si el químico no acepta el trabajo, con seguridad una persona muy inescrupulosa tomaría el trabajo y podría producir armas letales[49]. Desde una perspectiva utilitarista, lo que se busca es que la acción, en términos generales, produzca la mayor cantidad de beneficios para el mayor número de personas. Desde esta

[49] Es importante señalar que este ejemplo es usado por Williams como un caso de responsabilidad negativa (Williams, 1973a, p. 93ss): el utilitarismo, dice él, no solo responsabiliza al agente de lo que hizo, sino también de aquello que dejó de hacer. En el ejemplo del químico, este sería responsable de las consecuencias de aquello que voluntariamente dejó de hacer (*ibid.*, p. 97s).

perspectiva, se supone que el químico debería aceptar el trabajo porque es el curso de acción más deseable (para su familia) y también es el curso de acción que produce el menor daño posible (a la humanidad). Según el utilitarismo, el químico debería adecuar su acción a lo que produzca mayor bienestar general, sin importar los reparos éticos o políticos que tenga con respecto a la guerra química.

El segundo ejemplo propuesto por Williams es el de un botánico que durante su expedición por la selva se encuentra con un pelotón de soldados que está a punto de fusilar a veinte inocentes. El capitán del pelotón le ofrece al botánico la oportunidad de salvar a diecinueve de los acusados a cambio de que asesine a uno. En términos utilitaristas, es obvio que lo que debe hacer el botánico es asesinar a esta persona, pues el criterio moral utilitarista le impone buscar el mayor beneficio posible[50].

El lema utilitarista de "generar el mayor beneficio" es un proyecto de carácter impersonal e imparcial; no tiene en cuenta quién lleva a cabo la acción, no discrimina entre tipos de beneficio y tampoco importa qué tipo de acción haya que realizar para conseguir el beneficio óptimo. Por su carácter impersonal y abstracto, el utilitarismo desprecia la singularidad e individualidad de cada agente. Cualquier criterio de decisión basado en aspectos singulares y muy particulares de una persona es dejado de lado en el cálculo utilitarista. Si estos aspectos fueran tenidos en cuenta, el principio utilitarista se vería parcializado por criterios personales. De ahí que este modo de pensar se comprometa, ante todo, con un modelo de la agencia en el que los individuos son homogéneos, y sus intereses y deseos pueden medirse uniformemente en términos de placer y dolor.

El agente utilistarista, sin embargo, no puede únicamente tener el proyecto de maximinar la utilidad. Necesita otro tipo de proyectos menos homogéneos, más personales. Esto se justifica porque el agente utilitarista, cuando lleva a cabo el proyecto utilitarista, calcula el resultado deseable no porque se satisfaga ese proyecto en particular, sino porque *ese* proyecto satisface también otros "proyectos más básicos o de un orden más profundo" que tenga el agente como

[50] En este caso también se presenta la doctrina de la responsabilidad negativa: si acepta el ofrecimiento, el botánico es responsable de matar un hombre y de salvar a diecinueve; si no lo acepta, es responsable negativamente de haber matado a los veinte inocentes porque dejó de hacer voluntariamente algo que pudo haber salvado a casi todos ellos.

individuo (Williams, 1973a, p. 110). Que el químico acepte investigar para el desarrollo de armas no es deseable solo porque ese proyecto es el mejor según el cálculo utilitario, sino porque, por ejemplo, así cumple un deber patriota, o porque así logra sustentar a su familia, o porque si acepta puede boicotear la investigación. Lo que quiero decir es que el proyecto utilistarista general se sustenta, en parte, y de manera necesaria, por proyectos personales. Para que tenga sentido la idea de un "bien general" es necesario que haya bienes personales, bienes de "primer orden", de lo contrario el proyecto utilitarista "no tendría nada sobre lo que trabajar, y sería vacuo" (*ibid.*).

No existen, no pueden existir, agentes absolutamente impersonales e imparciales. El agente utilitarista, incluso el más comprometido que se pueda imaginar, además de ser un maximizador abstracto, es alguien que también tiene proyectos personales. Entre estos proyectos se encuentran deseos propios, deseos sobre la familia, los amigos y deseos que satisfacen necesidades vitales. También sus proyectos pueden ser de otra índole: intelectuales, estéticos, culturales, políticos o creativos (*cf.* Williams, 1973, p. 110). Incluso alguien puede estar comprometido con proyectos configurados por disposiciones generales como el odio a la injusticia a la crueldad o a la discriminación.

Existe entonces una diferencia entre los proyectos generales utilistaristas y los proyectos personales. Estos últimos no pueden estar al mismo nivel del proyecto imparcial utilitarista. Existen dos características de los proyectos personales que los distinguen de los proyectos generales utilitaristas. Primero, estos proyectos no necesariamente son satisfechos de acuerdo con el bienestar que produzcan, y de hecho poco o nada tienen que ver con el bienestar o el sacrificio que impliquen. Una persona puede estar dispuesta a perseguir su proyecto personal al margen del costo utilitario (para todos y para él mismo) que ello implique. Estos proyectos escapan al cálculo utilitarista de bienestar/malestar y no tienen que ver, en general, con la utilidad/placer o con el malestar que produzcan. Segundo, dentro de los proyectos personales, existen algunos tan básicos y fundamentales [*ground projects*], y son tan significativos e importantes para una persona, que ellos, como dice Williams, "sostienen su vida" (1981d, p. 11); con esto lo que quiero decir es que los proyectos fundamentales son aquellos que explican el interés que tiene una persona en seguir con vida. Por ahora dejo esta idea apenas enunciada; volveré sobre ella más adelante.

El fenómeno de los proyectos fundamentales da lugar al problema de la integridad personal. Este problema surge cuando las exigencias imparciales de la moralidad entran en conflicto con los proyectos más importantes y profundos de una persona (*ground projects*). El conflicto mismo representa una amenaza a la integridad personal. Para que una acción sea consistente con la integridad del agente, debe estar basada o cuando menos ser coherente con sus compromisos, proyectos y fines personales más importantes (*cf.* Rogerson, 1983, p. 462). Pero si una exigencia imparcial entra en tensión con un proyecto personal profundo, es absurdo pedirle a alguien que desconozca su propio proyecto y que actúe según el cálculo utilitarista. "Esto es enajenarle en un sentido real de sus acciones y de la fuente de su acción en sus propias convicciones" (Williams, 1973a, p. 116). En el caso del químico y del botánico, las reglas imparciales del utilitarismo les exigen actuar en contra de lo que pueden llegar a ser sus compromisos más profundos. Investigar para la guerra o asesinar a una persona puede suponer para ellos un sacrificio de sus compromisos éticos y personales más profundos. Ni la seguridad económica, ni la presión familiar –como en el caso del químico–, ni la posibilidad de salvar la vida de diecinueve personas –como en el caso del botánico–, son necesariamente razones para que ellos sacrifiquen su integridad. En palabras de Williams:

> Es absurdo pedir a tal hombre […] que debería apartarse de su propio proyecto y de su propia decisión y reconocer la decisión que requiere el cálculo utilitarista. Esto es enajenarle en un sentido real de sus acciones y de la fuente de su acción en sus propias convicciones. […] es descuidar la medida en la que *sus* acciones y *sus* decisiones han de ser vistas como las acciones y decisiones que surgen de los proyectos y actitudes con los que el agente está más fuertemente identificado. Esto es, en el sentido más literal, un ataque a su integridad (Williams, 1973a, pp. 116-7; énfasis en el original).

4.2 Integridad: carácter y autoindulgencia

El problema de la integridad consiste entonces en que una concepción imparcial de la moralidad representa una amenaza para la integridad del agente. Como he dicho, esta crítica está dirigida directamente contra del utilitarismo. Intentaré, sin embargo, mostrar que también es una crítica válida para el rigorismo moral kantiano. En lo que sigue, mi objetivo es, primero, desarrollar y explicar con precisión el problema de la integridad en los términos propuestos por Williams (§4.2)

y, posteriormente, explicar cómo este problema también afecta al rigorismo kantiano (§4.3).

Esta sección está dividida en dos partes. En la primera (§4.2.1), hago una caracterización positiva del concepto de integridad, lo cual supone interpretar la noción de integridad a partir de distintos textos de Williams, pues el autor no es sistemático ni explícito en el tratamiento de este concepto. En la segunda parte (§4.2.2), me detengo en el fenómeno de la auto-indulgencia moral. Este fenómeno debe interpretarse como una forma incorrecta en la que un agente llega a preocuparse por su integridad personal. En esta sección, entonces, espero caracterizar los sentidos positivo (qué es) y negativo (qué no es) de la integridad personal en los términos propuestos por Williams.

4.2.1 Integridad: identidad y carácter

El concepto de integridad, etimológicamente hablando, proviene del latín *integrîtas*: *-âtis* que se refiere a totalidad, virginidad, robustez. El vocablo propiamente deriva del adjetivo *integer,* que se refiere a estar intacto, entero, no tocado o alcanzado por un mal. *Integer* está compuesto del prefijo *in-* (no) y de la misma raíz del verbo *tangere* (tocar o alcanzar). "Cuando se aplica a objetos, la palabra [integridad] denota completitud [*wholeness*], unidad o pureza de una cosa" (Cox, et al., 2017). Cuando se refiere a personas, sin embargo, este concepto tiene una connotación distinta. En el uso cotidiano del término, decimos que una persona es íntegra especialmente para referirnos a una cualidad positiva o a una virtud de su carácter. Se piensa que una persona íntegra es alguien con la entereza o fortaleza para hacer lo que es correcto, y lo anterior suele estar asociado a la idea de 'hacer lo que es moral'.

En un contexto cotidiano, entonces, una persona exhibe o preserva su integridad justamente cuando actúa según lo que impone el deber moral. Williams, sin embargo, se separa radicalmente de esta idea. Para él, la amenaza a la integridad surge precisamente cuando la moralidad (imparcial) le exige al agente abandonar sus proyectos fundamentales. Esto significa no solamente que puede presentarse una tensión entre la integridad y la moralidad, sino que se está suponiendo una concepción muy específica de la moralidad, una que Williams asocia específicamente al utilitarismo (1973a; 1981e) y al kantismo (2006b, p. 174).

Cox *et al.* consideran que, en el contexto de la discusión sobre la integridad, la posición de Williams representa la Perspectiva de Identidad [*Identity View*] (Cox, et al., 2017, §2). Según esta interpretación, lo que Williams estaría afirmando es que la integridad personal debe entenderse en términos de la identidad del agente. En un sentido general, considero que esta interpretación va en la dirección correcta y la adopto parcialmente porque, como veremos en el capítulo 5, la principal crítica en contra de la posición de Williams asume esta interpretación. Sin embargo, entender la integridad en términos de identidad comporta un riesgo: el concepto de identidad ha sido asociado principalmente a Locke y al problema de la identidad personal, y esto puede llevar a confundir el problema de la identidad personal con el de la integridad. Sin embargo, se trata de ideas diferentes. Mientras Locke está interesado en encontrar criterios para identificar la entidad "persona" y con ello dar cuenta de cualidades o características que toda entidad-persona tendría que tener, Williams estaría interesado en la identidad de una persona singular (Cox *et. al.*, 2017). Por mi parte, considero que la idea que Williams tiene en mente, cuando piensa en integridad, no atiende a la identidad en sentido lockeano. Su preocupación remite más a la identidad de alguien particular, de alguien situado. Es por ello que el fenómeno que Williams ilustra se refiere a aquello que hace que una persona sea ella misma, es decir, se trata de la pregunta por '¿qué es ser uno mismo?'. Esta pregunta no necesariamente es incompatible con la pregunta de Locke, pero sí se trata de fenómenos diferentes: una cosa es preguntarse por la identidad de las personas, otra, preguntarse por la identidad de la persona que se es.

Para no confundir el problema de la identidad con la perspectiva de la identidad, me parece más apropiado pensar la integridad en términos del carácter de una persona. Con esta propuesta de lectura, no solo se evita la posible confusión con Locke, sino que además se captura con más precisión el sentido de la idea de Williams, pues la integridad personal se refiere a preservar (mantener completa la totalidad) 'aquello que la persona es', es decir, preservar aquellos rasgos psicológicos propios, singulares y caracter-ísticos de una persona particular. Me valgo de la etimología de la palabra 'carácter' que remite al griego *kharakter*, y que significa 'el que marca o graba'. De aquí proviene la idea de que el carácter es la marca o rasgo propio de una cosa, como un sello característico que determina lo que algo es. Cuando se extiende esta idea al campo filosófico, el carácter de una persona se refiere al conjunto de rasgos psicológicos, biológicos, culturales que tiene un individuo y que, en conjunto, configuran su forma o modo de ser. En este

sentido, mi interpretación de la posición de Williams es que la integridad personal ha de entenderse como la preservación o cuidado de la forma de ser propia de una persona. Si, como en el caso del biólogo o del químico, una persona se viera obligada a realizar acciones que van en contra de rasgos fundamentales de su carácter, básicamente ella estaría poniendo en riesgo aquello que es, estaría arriesgando lo que hace que sea quien es. Y esto, en mi opinión, es lo que significa que se amenace su integridad personal.

Hay que decir, en todo caso, que 'aquello que la persona es' es una idea que también puede capturarse en términos de identidad, pues preservar 'aquello que la persona es' es otra forma de decir que se preserva su identidad. Por esta razón, insisto, la interpretación de la integridad en términos de identidad me parece bien direccionada. Mi reparo en contra de Cox *et. al.* podría parecer, entonces, un asunto de precisión terminológica, ya que en realidad estamos apuntando a la misma idea: tanto la identidad de una persona como su carácter hacen referencia a los rasgos singulares, propios de una persona, y que constituyen lo que esta persona es. Y puedo aceptar que cuando Williams habla de una 'amenaza a la integridad personal' se está refiriendo a poner en riesgo su identidad o, más precisamente, su carácter.

Es verdad que la interpretación de Cox *et al.* es compatible con entender la integridad en términos del carácter, pues apuntan a la misma idea. Aun así, hay una diferencia que no es solo terminológica. Si bien, como mostraré a continuación, existe una interesante relación entre la identidad personal y el carácter de una persona, se trata de dos cosas diferentes.

a. Identidad y carácter

En su sentido más básico, el concepto de identidad hace referencia a la relación de una cosa, objeto o persona consigo mismo. En otras palabras, la identidad refiere a una forma de igualdad respecto a sí mismo. Este uso del concepto aplica para objetos físicos, objetos naturales, animales y en general para cualquier entidad —incluyendo personas— que se proyecte en el tiempo, pues para establecer la igualdad es necesario comparar la entidad en, al menos, dos momentos distintos. Si en esta comparación no se mantiene la igualdad (en el sentido relevante), entonces la identidad se pierde.

Para el caso de la identidad personal aplica la misma idea: para poder ser uno mismo es necesario preservar la identidad singular a lo largo del tiempo. Por ejemplo, para poder ser ella misma una persona debe preocuparse por mantener y preservar temporalmente el tipo de persona que es. Ella se ubica en un plano temporal y espera que el yo-actual esté vinculado positivamente con el yo-pasado y además debe estar preocupada porque ese vínculo se preserve en el futuro. Sin esta continuidad temporal la persona estaría desintegrada temporalmente, y pasaría a ser una sucesión de yoes dispersos en el tiempo.

¿Cómo establecer esta continuidad temporal entre el yo-pasado-presente-futuro? Esta continuidad, como afirma Williams en "Persons, Character and Morality", se explica a partir del hecho de que las personas tenemos un carácter*:* "una persona individual posee un conjunto de deseos, intereses o, como frecuentemente los llamaré, proyectos, los cuales ayudan a configurar un *carácter*" (Williams, 1981d, p. 5; énfasis en el original). El término *carácter*, para Williams, tiene la función de distanciarse de la idea de pensar a la persona como una sucesión de yoes: yoes pasados, presentes y futuros. Williams encuentra que este vocabulario de los yoes sucesivos es inadecuado porque no atiende al hecho de tener un carácter, es decir, a la idea de que los cambios en el tiempo que sufre una persona son "cambios en *su* carácter" (*ibid.* 9; énfasis en el original). Según Williams, la idea de continuidad en el carácter no puede ser capturada por una multiplicidad de yoes[51].

La idea es entonces la siguiente: si no existiera una continuidad en mi carácter, si existieran

[51] Esta idea de Williams debe entenderse en el contexto de su discusión con Derek Parfit sobre la identidad personal. Parfit (1987) defiende una versión lockeana de la identidad personal, a través de un psicologismo reductivista. Para este autor, lo que importa para la identidad es la 'relación R', es decir, la conexión y la continuidad psicológica entre 'sucesos' mentales. Ir más allá de la 'relación R' es comprometerse con la existencia de una entidad 'persona' que no puede ser descrita en términos reductivistas. Dentro de una visión reductivista, en cambio, no es necesario suponer ir más allá de la 'relación R' para establecer la identidad de una persona. La posición de Parfit implica que lo que debe importar con respecto a la identidad es la conexión y continuidad entre mi yo-presente y el (posible) yo-futuro. Si se pierde esta conexión, la persona perdería su identidad (p.e. como en los casos de la tele-transportación o los casos de fisión). Williams considera que centrarse en la 'relación R' convierte la identidad personal en un fenómeno "gradual" (Williams, 1981d, p. 6): la relación entre mi yo-presente y mi yo-futuro sería una cuestión de grados, al punto de que si no existe conexión o si la conexión es muy débil, mi yo-futuro sería otra persona. Para Williams, el psicologismo reductivista de Parfit, cuando se piensa en un escenario moral, no puede explicar la preocupación que tiene una persona por su propio futuro.

múltiples yoes, mi relación con mis proyectos futuros sería la misma relación que mi yo presente tendría con los proyectos de otra persona, pues mi yo presente asumiría que el yo futuro es alguien muy diferente al presente, un yo discontinuo. Pensemos en el ejemplo que Williams toma prestado de Parfit. Se trata de un noble ruso que está comprometido con la revolución y promete que en un futuro, cuando consiga su herencia, apoyará al ejército rebelde. El noble, además, le hace prometer a su esposa que le obligue a cumplir su promesa de apoyar la causa revolucionaria, sin importar lo que él piense en el futuro, incluso si para entonces ya no quisiera apoyar la revolución. En este ejemplo, el noble ruso se relacionaría con sus propios deseos y proyectos de la misma manera que se relacionaría con los deseos de otra persona; con lo cual perdería la continuidad de su carácter (Williams, 1981d, p. 9). Lo razonable, en cambio, es esperar que su compromiso con la revolución se mantenga en el tiempo; solo así diríamos que el noble está comprometido con la causa revolucionaria.

Tener carácter supone entonces un yo que está temporalmente vinculado. Y en eso consiste la interesante relación entre la identidad y el carácter: la pregunta por la identidad temporal, en el caso de las personas, se resuelve a partir de "tener un carácter". La relación de ambos conceptos radica en la temporalidad, pues tanto la identidad, como el carácter atienden a la continuidad de la persona en el tiempo. No obstante, para efectos de interpretar el concepto de integridad, lo que se gana con el carácter, que es dejado de lado con la identidad, es que esa continuidad temporal no puede desligarse de la preservación –en el tiempo– de esos rasgos que hacen parte del conjunto de deseos, intereses y proyectos que constituyen el carácter. Interpretar la integridad en términos de identidad sirve para resaltar el aspecto temporal de la integridad, algo que me parece esencialmente correcto. Solo que al pensar la integridad en téminos de carácter se tiene, no únicamente esta perspectiva temporal, sino que además se gana una perspectiva psicológica de la integridad.

b. 'Seguir hacia adelante' y deseos categóricos

La continuidad temporal es un elemento central para la integridad y se trata de un elemento capturado tanto en la identidad, como en el carácter. Por esta razón, tener un carácter va atado a una perspectiva temporal de la propia vida. Gracias al carácter, la persona puede entender su vida

dentro de un rango temporal y puede ofrecer una narración de su historia personal. Y, lo que me parece más importante, es en virtud del carácter que una persona se preocupa por el futuro de su propia vida. "El hecho de que una persona tenga cierto interés ahora en lo que hará o experimentará después requiere que posea ciertos deseos, proyectos o intereses ahora que se relacionen con tales acciones o sucesos posteriores […]" (Williams, 1981d, p. 8).

La preocupación por el futuro es una idea fundamental del hecho de tener carácter. Para Williams, si no existiera esta relación entre el presente y el futuro, no se explica cómo siquiera es posible para alguien seguir con la vida hacia adelante. En otras palabras, es en virtud de que tenemos una perspectiva prospectiva que podemos proyectar nuestros deseos, intereses, planes hacia el futuro (*cf.* Williams, 1981d, p. 11).

La preocupación por el futuro, sin embargo, tiene un límite inevitable: la muerte (*cf.* Williams, 1981d, p. 10), es decir, un estado en el que ya no es posible plantear la cuestión del futuro. Ella pone fin a nuestros deseos, intereses, planes, etc. La satisfacción de estos elementos depende, entonces, de que estemos vivos. Sin embargo, dice Williams, no todos los deseos están condicionados a que estemos vivos. Existe un tipo específico de deseos, los deseos categóricos (Williams, 1981d, p. 11), que son aquellos que nos dan una razón para estar vivos, aquellos que sostienen nuestra vida[52]. A diferencia de otros deseos, intereses o planes, los deseos categóricos no están condicionados a que estemos vivos, sino que son ellos los que dan soporte y una razón para estar vivos:

> Ciertos deseos dependen de que uno esté vivo, pero no todos los deseos pueden estar condicionados a este hecho. Pues es posible imaginar a una persona que considera el suicidio frente a un mal, y si decide seguir vivo será por cierto deseo que no puede operar condicionalmente a que este vivo, pues este deseo resuelve la cuestión de si seguir con

[52] La idea de deseos categóricos es una noción de raigambre humeana, y como tal, cumple un papel importante en la explicación de la motivación para la acción. Lamentablemente, en la literatura contemporánea no se le ha prestado mucha atención a este tipo de deseos, tal vez por su ascendencia humeana y por un cierto temor al irracionalismo. En todo caso, en el pensamiento de Williams este tipo de deseos cumple una función primordial.

vida o no. (Williams, 1981d, p. 11)

Los deseos categóricos hacen parte importante de nuestro sentido de tener un carácter. Williams apenas los menciona, pero se trata de deseos básicos que configuran la estructura volitiva del agente y establecen las condiciones para que él pueda seguir hacia adelante, para que pueda tener una perspectiva a futuro.

> El patrón de intereses, deseos y proyectos de uno no solamente proporciona la razón para tener interés en lo que sucede dentro del horizonte de su futuro, sino que también integra las condiciones para que tal futuro llegue a ser. (Williams, 1981d, p. 11)

c. Proyectos fundamentales y sentido [meaning]

Los deseos categóricos sustentan y dan forma a los proyectos fundamentales o proyectos base [*ground projects*[53]] de una persona. Un proyecto fundamental vendría a ser la formulación más o menos explícita, más o menos consciente[54], de un deseo categórico. No creo que sea necesario trazar una distinción precisa en este punto, basta con entender que en la vida tenemos fuerzas motivacionales tan profundas y propias (deseos categóricos) que se manifiestan la mayoría de las veces en un proyecto o propósito (proyecto fundamental). Mientras el deseo puede ser más amplio y hasta opaco, el proyecto supone el establecimiento de un fin o propósito. Por ejemplo, el deseo categórico de cuidar a mi hijo se manifiesta o expresa en mi proyecto fundamental de trabajar arduamente para ahorrar y así poder brindarle educación universitaria. El deseo categórico y el proyecto son dos caras de la misma moneda. Sin el deseo categórico, el proyecto carecería de fuerza motivacional —aquí es evidente la huella humeana de Williams—; y sin un proyecto establecido

[53] En "Persons, Character and Morality", el concepto de "proyecto" tiene una connotación existencial, y de hecho no sería equivocado rastrear su origen en el pensamiento de Heidegger (2012, §31) (*Sich entwurfen*: proyectarse o planearse a sí mismo; el *Dasein* es un "ser como proyecto") y de Sartre (2017) (*pro-jet:* el hombre es un proyecto inicial siempre abierto, es la manifestación de la libertad absoluta). Aunque Williams usa este concepto sin necesidad de vincularse con toda la carga teórica de estos autores, es verdad que los proyectos tienen un claro sello existencialista, pues son ellos los que fundamentan la base de nuestra razón para vivir, "un fundamento que le da sentido a nuestras vidas" (Rivera, 2007, p. 71).

[54] Para Williams, los deseos categóricos y los respectivos proyectos fundamentales que de allí surgen no tienen que ser completamente evidentes para la conciencia. Una prueba de esto es que una vida con sentido muchas veces es una vida cuyo sentido no se cuestiona, es decir, se puede vivir con sentido sin ser consciente de cuál sea el sentido de nuestra vida (Williams, 1981d, p. 12).

no es muy claro cómo podría expresarse o manifestarse el deseo categóricos, incluso para nosotros mismos.

Los proyectos fundamentales se distinguen de otro tipo de propósitos o planes por la importancia que estos tienen para las personas y por su papel en la formación del carácter (Rivera, 2007, p. 71). Según Williams, una característica propia de este tipo de proyectos es que "son a la vez más profundos y serios […] más individuales y vinculados [*permeated*] con el carácter que el deseo de las necesidades vitales" (Williams, 1973a, p. 111). La idea de proyecto fundamental no se agota en la racionalidad instrumental que pondera la relación medios y fines. El proyecto fundamental de una persona hace referencia, más bien, a aquellas cosas que son relativas al "'núcleo' o el 'centro' del grupo de preocupaciones con las que estructuramos nuestra agencia; ellas son importantes para nuestra identidad" (Rivera, 2007, p. 71). En palabras de Williams:

> Mis proyectos presentes representan la condición de mi existencia, en el sentido de que, a menos de que me vea impulsado [propelled] hacia adelante por el conato del deseo, proyecto e interés, no resulta claro por qué he de seguir adelante en absoluto (Williams, 1981d, p. 12).

Con esto, lo que Williams está diciendo, nada más y nada menos, es que los proyectos fundamentales confieren *sentido* [*meaning*] *a nuestras vidas* o "al menos representan una condición necesaria para la posibilidad de una vida con sentido: sin proyecto, no hay sentido" (Jenkins, 2006a, p. 33). Gracias a estos proyectos fundamentales, una persona tiene una dirección o un sentido que lo impulsa a seguir hacia adelante, que lo motiva a continuar con su vida para desarrollarlos: "los proyectos base de una persona proporcionan la fuerza motriz que lo impulsa hacia el futuro y le una razón para vivir" (Williams, 1981d, p. 13).

Si un agente fracasa rotundamente en la realización de sus proyectos fundamentales, ello significa que lo que dota de sentido a su existencia se ha perdido. Es en estos casos en los que cabría la expresión "¡No tiene sentido mi vida!". Sin embargo, algo como esto no implica que la única alternativa para quien así habla sea el suicidio (Williams, 1981d, p. 13). El agente ha de contar con otras cosas que lo aten al mundo. Aunque es verdad que un fracaso en uno de esos proyectos fundamentales sí es algo por lo que uno daría la vida, literalmente; hasta el punto de estar dispuesto

a morir para lograrlos (solo pensemos en el heroísmo o en los mártires). Williams cree que en general las personas no tienen un único proyecto: "más bien existe un nexo de proyectos, relacionados con sus condiciones de vida, y la pérdida de todos o la mayoría de ellos sería lo que elimina el sentido" (*ibid.*).

La idea de que los proyectos están relacionados con otros elementos de la vida de las personas es muy importante porque revela algo significativo de la naturaleza de los proyectos fundamentales. Comprometerse con un proyecto específico no significa que esto sea lo único que le importa a una persona. Dado que el proyecto se relaciona con una gran cantidad de elementos de la vida personal, la adherencia a un proyecto debería pensarse además como el compromiso con muchas otras cosas que también son importantes y valiosas para ella. De la misma manera que la punta del iceberg es solo la parte visible de algo mucho más profundo, asimismo el proyecto es solo lo que se manifiesta más o menos explícitamente; pero de fondo el proyecto está soportado por múltiples y varidos elementos como disposiciones, deseos, intereses, creencias, etc. Estos últimos dan forma a la compleja trama de nuestro carácter, y a su vez explican por qué el proyecto tiene tanto que ver con nuestra vida, y también explican por qué es tan costoso renunciar él.

Gracias a que los proyectos fundamentales se relacionan con otros elementos del carácter es que los primeros pueden tener tanto peso en la vida de una persona. En otras palabras, el valor que tiene el proyecto no es intrínseco, sino que depende de su relación con otro tipo de elementos del carácter que también son valiosos. Si el proyecto tuviera un valor intrínseco y fuera independiente de los otros aspectos constitutivos del carácter, entonces el fracaso del proyecto implicaría el derrumbamiento de una persona. Sin las bases, no habría forma de sostener el edificio. Esta explicación del valor relativo del proyecto fundamental explica por qué su fracaso no implica que la persona deba, por ejemplo, suicidarse o echarse a la pena. Aun cuando fracase aquello que nos da una razón para vivir y seguir adelante, hay en nuestra identidad y en nuestro carácter otros elementos que nos atan al mundo y que también son importantes para nosotros.

El carácter relacional de nuestra compleja vida práctica es central para explicar por qué los proyectos fundamentales son valiosos para nosotros. El valor de un proyecto no se mide por su grandeza, sus consecuencias o su trascendencia. Los proyectos tampoco tienen que ser

grandilocuentes o revolucionarios, basta con ellos sean la base "de muchos tipos de felicidad" (Williams, 1973a, p. 111). Tampoco existe un límite conceptual o una restricción racional o normativa sobre el contenido de estos proyectos: "Uno puede estar comprometido con cosas, tales como una persona, una causa, una institución, una carrera, al propio genio, o a la persecución del peligro" (*ibid.*). Además, no hay ninguna razón para pensar que los proyectos fundamentales no puedan ser altruistas o morales. Una persona puede estar comprometida con un bien social, una causa altruista o un compromiso moral; después de todo somos "individuos socializados", y gran parte de nuestros deseos y proyectos han sido "formulados por y desde adentro de disposiciones que constituyen un compromiso con la moralidad" (Williams, 1981d, p. 12).

Lo anterior, sin embargo, no significa que no existan ciertas condiciones para que algo cuente como el proyecto fundamental de alguien. Aunque en cierto sentido no haya límites conceptuales o normativos, desde el punto de vista del agente no todo puede contar como un proyecto fundamental. Para que esto sea así, hacen falta al menos tres condiciones indispensables. La primera de ellas es que el proyecto debe estar vinculado a cosas que son importantes para la persona (en §6.2.2.b desarrollaré el concepto de importancia). Es cierto que el proyecto está vinculado a una compleja trama de elementos del carácter, pero habría que agregar algo más: algunos de los elementos que se vinculan al proyecto deben ser importantes para la persona, de lo contrario, el proyecto fundamental no sería valioso ni importante para ella. Si una persona afirma que "no puede imaginar su vida sin fumar", parecería que ella tiene el proyecto fundamental de fumar[55]. Y dado que no hay límite conceptual o normativo para los proyectos, tendríamos que aceptar esta conclusión. Sin embargo, la conclusión no tiene en cuenta la conexión que debe existir entre el proyecto y otros elementos de la vida y del carácter. En otras palabras, el proyecto fundamental no tiene valor por sí mismo, sino en relación con otras cosas que tienen valor. Por ejemplo, 'fumar por fumar' no puede ser un proyecto en sí mismo, a menos que esté ligado a otros elementos significativos y valiosos de la vida (recuerdos, asociaciones con algo o alguien, cierto tipo de sensación corporal, cierto tipo de estado de ánimo, etc.). De otra manera, el enunciado "mi vida es inimaginable sin fumar" carecería de sentido, a menos que pueda ser vinculado a una red de elementos que sean importantes para quien así habla[56].

[55] Agradezco este ejemplo a Víctor Verdejo.

[56] Esta idea está directamente relacionada con el concepto de '*disciplina narrativa*' (Corbí, 2012, p.100ss). Para

Una segunda condición es que la persona debe estar identificada con su proyecto fundamental[57]. Una persona se identifica con sus proyectos cuando se mantiene incondicionalmente fiel [*steadfastly true*] a ellos (Cox, et al., 2017, §2). En pocas palabras, tener un proyecto fundamental que sea importante en la vida implica identificarse con él. Los proyectos fundamentales son una expresión muy propia y singular de la identidad y del carácter de una persona. Ciframos nuestra identidad en gran medida por la identificación que hacemos con estos proyectos. De hecho, lograr una correspondencia entre la vida, la acción y los proyectos es una de las formas más claras de reafirmar quién es uno. Si alguien se identifica con un proyecto, se esperaría que su fidelidad a él sea estable en el tiempo, es decir, que el yo pasado, presente y futuro se identifiquen con el mismo proyecto.

---

entender este concepto me remito a una tesis de Isaiah Berlin sobre las condiciones de inteligibilidad de las creencias y prácticas de una sociedad diferente a la nuestra (Berlin, 1958, p. 11s). Si esta sociedad adora un trozo de madera esto solo puede ser inteligible si existe una asociación entre la madera y "alguna otra característica o situación que reconozcamos como importante o sagrada" (Corbí, 2012, p. 98). Siguiendo a Berlin, Corbí considera que para identificar una acción u objeto como algo con valor V, es necesario "aportar una historia [*story*] que conecte tal acción u objeto con alguna característica que soporte la relación relevante con el valor V" (*ibid.*). De esta manera, no cualquier cosa puede ser objeto de valor. Solo aquello que tenga una disciplina narrativa, es decir, solo algo que haga parte de una narración que exhiba la conexión que existe entre las cosas que tienen valor, puede ser objeto de valor para nosotros. Por ejemplo, adorar un trozo de madera, fumar, o escalar montañas son actividades que, por sí solas, no exhiben valor.

[57] El concepto de *identificación* fue introducido por Harry Frankfurt (1988) en el contexto de la discusión sobre el concepto de *persona*. Frankfurt considera, en contra de Strawson (1959, p. 101-2), que el concepto de persona no se agota en sus características racionales (o en su mentalidad, como lo llama Strawson) (*cf.* Chandler, 2004). En cambio, Frankfurt considera que la mentalidad no es suficiente para dotar de personalidad a un ser humano. No solo tenemos creencias y deseos (actitudes proposicionales), sino que además tenemos actitudes [*attitudes*] sobre estas actitudes (*ibid.* p. 16). En otras palabras, reflexionamos sobre nuestras actitudes. Esto da lugar a la conocida distinción entre deseos de primer y segundo orden. Es un hecho que tenemos deseos de primer orden (deseo fumar, por ejemplo), pero nuestra disposición ante ellos da lugar a deseos de segundo orden (deseo no desear fumar). Para Frankfurt, lo característico de una persona es la disposición de su voluntad, no su mentalidad (Frankfurt, 1988, p. 17). Un ser humano adquiere personalidad [*personhood*] cuando se identifica con su voluntad. Y un ser humano se identifica con su voluntad cuando existe una correspondencia entre sus deseos de primero y segundo orden. En esta investigación no me adhiero a la postura de Frankfurt, aunque muchas de sus conclusiones son consistentes con mi análisis. Mi dificultad con Frankfurt es que resulta difícil entender cómo la noción de carácter se inserta en la estructura de deseos de primer y segundo orden. O bien el carácter se constituye exclusivamente a partir de deseos de segundo orden, en cuyo caso el carácter sería un fenómeno excesivamente autoconsciente; o bien el carácter sería una mezcla de deseos de primer y segundo orden, algo que me parece más plausible, pero entonces hace obsoleta la distinción entre deseos de primer y segundo orden.

La identificación garantiza cierta continuidad, no solo psicológica, sino volitiva: hay estabilidad de la voluntad gracias a que existe un proyecto fundamental. Cuando alguien alcanza un compromiso estable con sus proyectos y cuando moldea su vida de acuerdo con ellos, definitivamente ha logrado una identidad propia y adquiere un sentido de integridad: mi vida, mis proyectos y mis acciones reflejan el tipo de persona que soy, –incluso, reflejan el tipo de persona que quiero ser. Es en este sentido que quien vive de acuerdo con sus proyectos fundamentales manifiesta quién es, o en otras palabras, manifiesta su carácter y su identidad.

La tercera característica de los proyectos fundamentales es que no están enteramente bajo el control voluntario de la persona. Esto se explica, en parte, por el vínculo entre el proyecto y los demás elementos del carácter. Y es que si este vínculo existe, evidentemente tendría que haber cierta coherencia entre los deseos categóricos y los proyectos fundamentales de una persona con otros elementos de su carácter. No puedo desear categóricamente cuidar a mis hijos y embarcarme en proyectos que obstaculicen este deseo categórico. Eso implicaría una contradicción entre la estructura de mi voluntad y mis proyectos. La relación entre los proyectos y el carácter de una persona muestran que la persona tiene un marco previo a partir del cual configura sus proyectos. Los proyectos no surgen espontáneamente de la nada sino que están atados y soportados por elementos del carácter. Por esto es que el compromiso de una persona con su proyecto no es una decisión sin condiciones, no es una elección desde la nada; más bien, es un compromiso que expresa rasgos característicos de su carácter. De manera que la persona puede tener control sobre cómo llevar a cabo su proyecto, sobre su alcance, sobre la duración, e incluso, puede decidir no seguir con el proyecto; pero no puede elegir tener justamente ese proyecto. Este es el sentido en el que el proyecto personal no está enteramente bajo su control.

La falta de control voluntario sobre los proyectos remite a algo mucho más profundo de lo dicho hasta acá. Los proyectos fundamentales no están controlados por la voluntad del agente, más bien, ellos obedecen a una necesidad práctica[58]. Más adelante, en §6.2.2.c, desarrollaré este concepto, pero preliminarmente cabe decir que los deseos categóricos imponen una necesidad sobre la persona contra la que no se puede —o es muy difícil— luchar. La fuerza de esta necesidad radica

[58] Agradezco a Josep Corbí la idea pensar que los proyectos fundamentales obedecen a un tipo especial de necesidad práctica.

en que está apoyada en deseos categóricos. El impulso de su deseo empuja a alguien a actuar para satisfacer ese deseo. Pero no se trata de un impulso cualquiera, se trata de una fuerza que se convierte en una necesidad, en algo que 'no puedo dejar de hacer'. Los deseos categóricos y los proyectos fundamentales comportan este tipo de necesidad práctica. Esto explica por qué se escapan a un control voluntario y por qué uno se resiste a abandonarlos: ¿cómo hacerle contrapeso a algo que simplemente 'no puedo dejar de hacer'? Volveré sobre esta idea más adelante.

d. Amenaza a la integridad: proyectos fundamentales y carácter

El análisis anterior es útil para entender mejor en qué consiste la amenaza a la integridad que preocupa a Williams. En las páginas anteriores, caractericé la integridad como la preservación o el cuidado de los rasgos del carácter. Además, mostré que uno de los elementos del carácter son los proyectos fundamentales de una persona que dotan de sentido su vida. Es importante entender que hay una distinción entre el carácter y los proyectos fundamentales de una persona, aunque ambas nociones están estrechamente relacionadas. El carácter está constituido por el conjunto de sus disposiciones, deseos, intenciones, fines, vivencias, biografía personal, deseos categóricos, proyectos fundamentales, etc. Por tanto, el carácter no se reduce a sus proyectos fundamentales.

Esta diferencia entre carácter y proyectos fundamentales me lleva a diferenciar una amenaza a la integridad cuando están en riesgo los proyectos fundamentales, de una amenaza a la integridad cuando está en peligro el carácter. Esta distinción es apenas conceptual y dudo mucho que en la vida práctica de un ser humano tal distinción pueda ser trazada. Digo esto, en primer lugar, porque, como acabo de mencionar, los proyectos fundamentales no son independientes de muchos otros elementos del carácter; y segundo, porque muchas disposiciones del carácter surgen de los proyectos fundamentales y los deseos categóricos que una persona tiene, de manera que amenazar al proyecto supone amenazar también las disposiciones que dependen de este (Jenkins, 2006a, p. 38). Por estas razones, la amenaza a la integridad difícilmente tiene un objeto concreto dentro del vasto universo de la psicología de las personas. Esto es una muestra de la naturaleza compleja y rica de nuestra vida práctica que se expresa en nuestro carácter, en nuestras disposiciones y en nuestros deseos categóricos y proyectos más profundos y propios. Seguramente sería un error pretender dar cuenta precisa del objeto de la integridad personal, pues hacerlo no reflejaría la

complejidad de nuestras vidas y de nuestras acciones. Como bien señala Williams, las reflexiones sobre la integridad pueden ser confusas, no es muy claro "hacia qué estaría dirigida la integridad" (Williams, 1995, p. 212; énfasis en el original). Lo importante, por ahora, es advertir que el carácter es un conjunto de muchos elementos que están relacionados entre sí. Muchos de estos elementos son contingentes; otros no tanto. "[A]lgunas cosas son en algún sentido lo que eres, o expresan lo que eres, y otras no" (Jeffries, 2002, p. 4; citado en Guignon, 2004, p. viii). Lo cierto es que todos estos elementos se vinculan entre sí de una forma específica que depende de las condiciones de vida de cada uno (Williams, 1981d, p. 13).

Empiezo caracterizando el caso en el que se ponen en peligro los proyectos fundamentales. Como he dicho, estos proyectos fundamentales tienen una carga existencial. Abandonar o poner en peligro los proyectos fundamentales equivale a renunciar también a aquello con lo que alguien se identifica fielmente, y por ello, en un sentido muy relevante, pone en peligro aquello que lo hace seguir adelante. En mi interpretación, este es el sentido en el que a Williams le interesa hablar de integridad. Para él, tal y como lo expresó en "A Critique of Utilitarianism", una persona que en nombre de reglas morales imparciales se vea obligada a actuar en contra de sus proyectos fundamentales es víctima de una amenaza a su integridad. Cuando una exigencia moral entra en tensión con un proyecto personal fundamental, es absurdo pedirle a un ser humano que se desvincule de él. "Esto es enajenarle en un sentido real de sus acciones y de la fuente de su acción en sus propias convicciones" (Williams, 1973a, p. 116).

El ejemplo del pintor Paul Gauguin ilustra cómo el rechazo a los proyectos fundamentales constituye una amenaza a la integridad (Williams, 1981c)[59]. Gauguin decide abandonar a su familia, su trabajo y su vida en Francia para marcharse a pintar a Tahití. Esta decisión es tomada por él bajo la incertidumbre acerca de si su proyecto sería o no exitoso; la toma en contra de su salud, de su familia y dispuesto a pasar por penurias económicas. Durante su vida Gauguin nunca obtuvo ni fama ni reconocimiento; murió precariamente, enfermo y aislado y, sin embargo,

[59] Williams (1981c, p. 22ss) usa el ejemplo de Gauguin en el contexto de la discusión sobre la fortuna moral, más específicamente, lo usa para señalar que la deliberación de un agente (su justificación ante otros y ante sí mismo) en este caso esretrospectiva, pues la decisión depende, en gran medida, del éxito del proyecto. Sin embargo, el éxito solo es alcanzable después de que la decisión ha sido tomada y además también depende de la fortuna, es decir, de varios elementos que están fuera del alcance del agente.

consciente del peso de la decisión que tomó. Él se desvinculó de las exigencias sociales y morales y, en cambio, estuvo profundamente comprometido con su proyecto de pintar. Para su caso, podría darse uno a la tarea de evaluar si su decisión de marcharse a Tahití está o no justificada moralmente. Como lo veo, esto resulta bastante difícil y por suerte no es relevante para ilustrar el argumento de Williams. Lo que me interesa resaltar del ejemplo es, en primer lugar, que lo que determina la identidad y la integridad de un agente bien puede ir en contra de las demandas morales; y segundo, que no es obvio que la moralidad sea y deba ser el criterio último para decidir sobre lo qué hacemos con nuestra vida.

Otra forma de entender la amenaza a la integridad es verla a la luz de los casos en los que las demandas morales comportan un riesgo para el carácter de una persona y no solamente para sus proyectos fundamentales. En mi lectura, los ejemplos del químico y del botánico que ofrece Williams nos ayudan a entender los casos en los que la integridad se ve amenazada por la vía de una amenaza al carácter. Igualmente, pienso que el caso del asesino en la puerta es un escenario en el que también hay una amenaza al carácter de la persona que lo enfrenta. El químico se niega a investigar para el desarrollo de armas químicas, el botánico se niega a asesinar a un hombre, y muchos estamos dispuestos a negarnos a ser veraces con el asesino. A diferencia del caso de Gauguin, es difícil pensar que el químico tenga un proyecto fundamental negativo del tipo "no investigaré para el desarrollo de armas químicas" o que el botánico tenga un proyecto como "no asesinaré a ningún hombre" y difícilmente tenemos un proyecto como "mentir a los asesinos". Más bien, la negativa para llevar a cabo estas acciones se explica en estos casos por el carácter, es decir, por el tipo de disposiciones morales que tienen el químico, el botánico y nosotros. El químico rechaza el trabajo porque está motivado por una "amalgama de reverencia y honor" (Jenkins, 2006a, p. 38), y yo añadiría, por una disposición a no dañar a otras personas. El botánico, por su parte, no asesina al hombre por una amalgama de respeto a la vida, honor o algún sentido de justicia que le hace pensar que su posible víctima es inocente. Y lo mismo podría decirse de la persona que miente al asesino, pues su acción está motivada por un interés en cuidar a su amigo y en salvar una vida inocente. Como explicaré en §6.2.2.c, el rechazo a cometer estos actos se explica porque estamos impedidos por necesidad práctica a realizar actos que nos resultan impensables.

En estos tres casos, las personas se rehúsan a actuar como lo exige la norma moral. En el caso del

químico y del botánico, su rechazo va en contra del principio utilitarista; en el caso de quien miente al asesino, lo hace en contra del deber perfecto de la veracidad. Todos ellos rehúsan la norma moral debido al tipo de disposiciones del carácter de cada uno de ellos. Como vimos, al actuar en contra del carácter se atenta también contra la identidad. De actuar como se les exige, la relación que estas personas tendrían con las acciones que reprueban, sería la misma relación que tendrían con las acciones realizadas por otra persona con otro carácter y con disposiciones distintas. Actuar en contra del carácter es, en una palabra, una forma de actuar en contra de sí mismo.

e. Muerte práctica

Distinguir la amenaza a la integridad cuando está en peligro el proyecto fundamental del caso en el que está en peligro el carácter es importante para señalar que lo que se amenaza, en cada caso, es distinto. Ambos casos representan una amenaza a la integridad, solo que esta amenaza puede manifestarse de distinta manera. En el caso de los proyectos fundamentales, la amenaza a la integridad se da porque sacrificar un proyecto fundamental significa poner en peligro la razón que tiene la persona para seguir adelante, es decir, la razón que tiene para vivir. En el caso del carácter, la amenaza a la integridad se da a través de la ruptura temporal entre el yo antiguo y el yo futuro. Hay, por tanto, una ruptura en la identidad de la persona. Específicamente, si una persona renuncia a su carácter, estaría aniquilando la idea de que su yo futuro sea el mismo que su yo presente, estaría actuando como si fuera otra persona.

Considero que una forma de entender lo que está en juego cuando se presenta una amenaza a la integridad, ya sea en el caso del carácter o del proyecto fundamental, es pensando en el sentido del término *muerte práctica*: "hay algo que es semejante o incluso peor que la muerte: la muerte práctica, es decir, la posibilidad de dejar de ser uno mismo" (Korsgaard, 1996, p. 18).

Evidentemente, la muerte práctica no equivale a una muerte biológica. Pero sí existe una semejanza: y es que la muerte práctica significa el fin de una forma específica de vida. La muerte práctica, al igual que la muerte biológica, significa no vivir más la antigua vida. Aunque exista una continuidad en el sentido biológico, la persona que ha abandonado sus proyectos o rasgos de su carácter, ha renunciado también a la continuidad volitiva, y es por esta razón ha que dejado de ser

ella misma. Más bien, quien ha renunciado a su carácter o a su proyecto fundamental se ha desconectado de su yo antiguo.

Realizar una acción que vaya en contra del proyecto fundamental o del carácter puede conducir a que el yo antiguo no se reconozca como el autor de esta acción. Desde luego, la persona que así actúa puede asumir la responsabilidad causal y moral de lo que hace, solo que experimenta que es una acción que no se corresponde con la persona que ella es.

La preocupación que una persona exprese por su integridad está atravesada por la preocupación por el futuro como siendo suyo. Las acciones que realiza y que va a realizar han de estar vinculadas y ser coherentes con su carácter y sus proyectos. De no ser así, acontece esta ruptura del yo. Por esta razón, la preocupación por la integridad tiene que ver con una preocupación por no padecer muerte práctica.

4.2.2 Autoindulgencia moral

Hasta este punto del análisis, he tratado de ofrecer una caracterización positiva de la integridad. He interpretado este concepto en términos de la identidad y del carácter de una persona. Para ofrecer una caracterización más completa de la integridad habría que explicar además el sentido en el que, psicológicamente al menos, una persona está genuinamente preocupada por su integridad. Esta explicación, como veremos, cuenta como una advertencia sobre una forma inadecuada de pensar la integridad.

Williams indica que una persona preocupada genuinamente por su integridad no puede ser alguien que tenga como motivación explícita "preservaré mi integridad" (Williams, 1981e, p. 40). Como he tratado de señalar anteriormente, una persona genuinamente preocupada por su integridad está preocupada por no renunciar a su carácter, a sus disposiciones o a sus proyectos, porque son estos elementos los que le otorgan y sostienen su sentido de integridad. Pero esta preocupación genuina no equivale a estar preocupado por la integridad misma; en otras palabras, el objeto de la integridad no puede ser la integridad en sí misma. Si así fuera, aquella persona no estaría preocupada por su integridad, sino por su auto-imagen: la de ser-una-persona-íntegra. Y esta es

una forma incorrecta de vivir la integridad. Ampliaré esta idea con el fenómeno de la autoindulgencia moral.

Williams (1981e) advierte sobre el fenómeno de la autoindulgencia moral y su relación negativa con la integridad. Este fenómeno surge en escenarios en los que existen fuertes razones de índole moral para hacer algo que nos resulta detestable [*detestable*][60] (Williams, 1981e, p. 42). En el caso del utilitarismo, si el cálculo utilitarista nos exige realizar una acción *p*, no cuenta como una razón en contra de *p* que se trate de una acción detestable. El ejemplo del botánico muestra que existen buenas razones utilitaristas para asesinar a una persona a cambio de salvar a otras diecinueve. El caso del químico muestra que existen buenas razones utilitaristas para aceptar el trabajo en el laboratorio de armas químicas. Sin embargo, estas acciones tienen la característica de que resultan detestables [*detestable*] para ellos.

Para el utilitarismo, que una acción le resulte a alguien detestable no cuenta como una razón para no llevarla a cabo, si lo que está en juego es el principio de utilidad. Pero, al margen del utilitarismo, que una acción sea detestable puede contar como una razón válida para negarse a realizarla. Cuando alguien considera que una acción como producir armas químicas o asesinar a un inocente es detestable, podría resistirse llevarlas a cabo, por más que estos sean los cursos de acción exigidos por el principio de utilidad. Y es verdad que el rechazo del botánico o del químico a actuar en

[60] En mi interpretación, Williams usa la expresión "acción detestable" [*detestable action*] para referirse a acciones que resultan impermisibles para el agente, pero no necesariamente desde una perspectiva moral. Una acción detestable no equivale a una acción "moralmente prohibida" ni a una acción "contraria al deber moral" justamente porque lo que cuenta como detestable no necesariamente depende de que la acción cumpla reglas o principios morales. Una característica de las acciones detestables es que la persona las encuentra moralmente desagradables [*morally distasteful*] (Williams, 1981e, p. 40). En "Utilitarianism and self-indulgece" (1981e), Williams parece usar los adjetivos detestable [*detestable*] y desagradable [*distasteful*] como sinónimos para referirse a estas acciones impermisibles que demanda el utilitarismo. Sin embargo, hablar de desagrado moral [*moral distaste*] introduce el elemento humeano del gusto [*taste*] en el escenario; el problema es que no se explica el peso normativo de este desagrado. ¿Acaso el desagrado moral es suficiente contrapeso normativo frente a la deliberación moral utilitarista? Para evitar este cuestionamiento, solo me comprometo con la categoría *detestable* para caracterizar aquellas acciones que, como en los casos del botánico o del químico, demanda el utilitarismo y que, sin embargo, van "en contra de los compromisos personales más fuertes" (Williams, 1981e, p. 40). La categoría *detestable*, como la interpreto, adquiere su fuente normativa de la ética (atenta contra valores, ideales) y de razones personales (proyectos, carácter) y no con con el hecho de que la acción contradiga una regla moral.

favor de ese principio es una expresión del tipo de carácter que ellos tienen; nos dice algo sobre los valores a los que se adhieren: si ellos no siguen el curso de acción utilitarista, sabríamos que son sensibles a razones no-utilitarias. Aun así, también existen razones no-utilitarias que, sin embargo, son muy cuestionables:

> Alguien que conscientemente sigue el curso anti-utilitario en tal caso podría verse sujeto a ser acusado de preocuparse por su propia integridad o pureza o virtud a expensas de otros. Para utilizar una frase como título general, aunque no siempre resulte la mejor frase, llamemos a esto la acusación de autoindulgencia moral (Williams, 1981e, p. 64).

La autoindulgencia moral es un tipo específico de motivación que surge cuando alguien se resiste a realizar una acción detestable porque el objeto de su preocupación es la pureza de su virtud o la preservación de su integridad. Por ejemplo, si el botánico se niega a asesinar a una persona por autoindulgencia moral entonces "lo que le preocupa a él no son tanto las demás personas, sino el hecho de que él mismo se preocupe por otros" (Williams, 1981e, p. 45). El autoindulgente está entonces preocupado por verse a sí mismo como una persona íntegra, mientras que su preocupación por otras personas es secundaria, accidental o incluso inexistente.

El fenómeno de la autoindulgencia moral es una forma deformada de reflexividad en la que el agente sobrevalora su integridad o su virtud, al punto de que la convierte en la única o en la principal motivación para hacer o dejar de hacer algo. La autoindulgencia moral es un fenómeno práctico que demuestra una comprensión incorrecta de la integridad personal.

Con frecuencia, el motivo que se impone un agente autoindulgente está dado desde la perspectiva de un agente ideal; es decir, de aquel que actúa sobre la base de un modelo prescriptivo de lo que haría una persona virtuosa o integrada. Sin embargo, mientras hace esto, él pierde de vista su propio carácter y sus propias disposiciones. No es lo mismo realizar un acto generoso porque se tiene la disposición a hacerlo, que realizarlo porque "es lo que haría una persona generosa".

Basado en lo anterior, Williams considera que la falla moral de la autoindulgencia radica justamente en que "implica una inversión en una división que es fundamental para cualquier vida moral o incluso cuerda, entre la preocupación por sí mismo y la preocupación por los demás;

implica una dirección errada no sólo de la atención […], sino genuinamente de la preocupación [...]" (Williams, 1981e, p. 47). En otras palabras, en el agente autoindulgente la preocupación por los demás se reduce a –se convierte en– la expresión de una preocupación por sí mismo, lo cual es un absurdo. De acá, concluye Williams que la integridad no es una virtud en absoluto (Williams, 1981e, p. 49). Aunque ello no signifique que la integridad no sea un elemento favorable o positivo en nuestras vidas. El punto es, más bien, que la integridad "no se relaciona con la motivación de la misma manera en que lo hacen otras virtudes" (*ibid.*)

> [La integridad] [n]o es una motivación que en sí produzca motivaciones, como la generosidad o la benevolencia; tampoco es una virtud de las a veces llamadas virtudes "ejecutivas", que en sí no producen una motivación característica, pero que son necesarias para la relación entre uno mismo y el mundo que le permite a uno actuar a partir de motivaciones deseables en formas deseables, el tipo que incluye la valentía o el autocontrol. (Williams, 1981e, p. 49)

Si la integridad se considerara, por sí sola, una virtud, no escaparía de la sombra de la autoindulgencia moral. Si la integridad fuera considerada una motivación parece muy difícil describirla sin caer en la deformación reflexiva que da lugar a la autoindulgencia. Lo que sugiere Williams, y estoy de acuerdo con él, es que alguien que esté genuinamente preocupado por su integridad no debe preocuparse por la integridad misma, sino por el contenido de aquello que constituye su integridad: su identidad, su carácter y sus proyectos. Alguien que muestra integridad actúa según las disposiciones y motivaciones que son más profundamente suyas y con las que más se identifica (*cf.* Williams, 1981e, p. 49), es decir, aquellas que emanan de su carácter o que forman sus proyectos fundamentales.

4.3 El problema de la integridad y la concepción del yo dividido

No es difícil entender de qué manera el rigorismo supone una amenaza para la integridad, pues —al igual que sucede con el utilitarismo— también en este caso alguien puede verse obligado por el deber racional a abandonar disposiciones de su carácter o sus proyectos fundamentales, y con ellos pone en riesgo su identidad y, por tanto, también su integridad. Williams afirma esta idea de modo más contundente:

> Puede llegar un punto en que sea bastante irrazonable que la persona, en nombre del buen orden imparcial del mundo de los agente morales, renuncie a algo que es una condición para que tenga interés en estar en el mundo. (Williams, 1981d, p. 13)

Esta crítica impacta directamente al modelo kantiano de la agencia. El problema de la integridad, lejos de ser secundario, es intrínseco al modelo de la agencia sobre la que se fundamenta la filosofía kantiana. Aunque el problema de la integridad, en la versión de Williams, concierna al utilitarismo, lo cierto es que el rigorismo kantiano también representa una amenaza para la integridad. Básicamente, el kantismo, al igual que el utilitarismo, tienen una comprensión impersonal de la moralidad, con lo cual abren la posibilidad de que las exigencias morales puedan ir en contra de la integridad de una persona. Lo que quisiera resaltar es que el problema de la integridad, en el contexto de la moral kantiana, está atado y se desprende de la concepción del yo dividido.

A primera vista, la concepción del yo dividido representa la imagen de que el agente kantiano se encuentra estructuralmente dividido. La concepción de un yo dividido es evidentemente una forma de fracturar la agencia. Y bajo este modelo, el agente kantiano ha de auto-comprenderse, ante todo, como un ser racional que debe controlar los aspectos no racionales de su vida. Semejante escenario nos deja con una visión esquizofrénica del agente (Stocker, 1976, p. 454ss). Si el agente se encuentra estructuralmente dividido, esa división parecería representar un problema para la integridad personal, pues un agente dividido sería, por eso mismo, un agente desintegrado. Como se puede advertir, esta intepretación solo tendría sentido bajo el supuesto de que la integridad atiende o se centra en un tipo de unidad o totalidad (algo que no es del todo alejado de la etimología del concepto y de un uso cotidiano del término). En el siguiente capítulo abordaré esta perspectiva, y mostraré de qué manera la concepción del yo dividido sería un problema para una intepretación que intente articular la integridad en términos de unidad. Esta lectura, sin embargo, no se corresponde con la interpretación que he propuesto en este capítulo, pues la idea de integridad, como se ha desarrollado en esta investigación –y como creo que la entiende Williams– ha de ser entendida en términos de carácter, y no en términos de unidad.

Acusar a la concepción del yo dividido de escindir al agente no es suficiente para derivar el problema de la integridad personal. Hay que explicar, además, por qué la des-unificación del yo

kantiano representa una amenaza al carácter y a los proyectos fundamentales de las personas. ¿Cuál es, en últimas, la amenaza de la concepción del yo dividido a la integridad del agente? Creo que puedo articular dos respuestas distintas a esta pregunta. La primera muestra la tensión entre este modelo de la agencia y la integridad; la segunda, sería una crítica más fuerte que muestra que la concepción del yo dividido hace ininteligible la preocupación de un agente por su integridad.

Pensemos primero cómo encajan las categorías de la integridad (carácter, deseos categóricos, proyectos fundamentales) en el modelo kantiano de la agencia. El yo kantiano está dividido en una parte racional y otra parte empírica-fenoménica (§3.4). La naturaleza de ambas partes es muy divergente. El yo racional no tiene rasgos disintitivos, ni tiene cualidades particulares ni idiosincráticas que lo distingan del resto de agentes racionales. En una palabra, el yo racional es impersonal:

> Solo puede ser una voluntad autónoma la que incorpora lo que Kant llama voluntad "pura". En otras palabras, debe ajustarse a las exigencias de una voluntad que es indiferente a todos los intereses personales: está por completo vacía de motivos, preferencias y deseos empíricos. […] Su pureza [la de la voluntad] consiste precisamente en el hecho de no verse afectada en absoluto por ninguno de los rasgos contingentes que hacen distintivos a los individuos y caracterizan sus identidades específicas. […] *La voluntad pura no tiene individualidad alguna. Es idéntica a todos y sus voliciones son por doquier exactamente las mismas. En otras palabras, la voluntad pura es del todo impersonal. Los mandatos que emite no son emitidos por nadie en particular*. (Frankfurt, 2007, p. 209s; énfasis añadido).

Si el yo racional carece de individualidad y de identidad personal, ¿dónde habría que ubicar, dentro de la concepción del yo dividido, al carácter, a los deseos categóricos y a los proyectos fundamentales? Por fuerza, y en contraste con el yo racional, no cabe más remedio que ubicar estos estos elementos en la parte fenoménica. La parte racional no tiene cabida para el carácter ni para los proyectos fundamentales, estos fenómenos no son característicos de un yo racional impersonal y puro.

Si esta caracterización es correcta, resulta que, desde la perspectiva del yo racional, no es posible que alguien esté genuinamente preocupado por su integridad. ¿Cómo podría entonces preocuparle

su carácter o sus proyectos fundamentales? O para ser más precisos, ¿cómo podrían preocuparle más que las exigencias normativas-morales del yo racional?

La conclusión que derivo de todo esto es la siguiente: la concepción del yo dividido representa un problema a la integridad personal. Dado que el yo racional tiene jerarquía normativa por encima de la parte no racional, resulta cierta la acusación de Williams de que en "nombre del buen orden imparcial del mundo de los agentes morales" la persona tiene que renunciar a su carácter o a sus proyectos fundamentales. En este sentido, el agente kantiano, en nombre de las exigencias normativas del yo racional, amenaza su propia integridad personal.

La anterior conclusión solo demuestra la tensión entre la concepción del yo dividido y la integridad. Esta primera crítica, como la acabo de planetar, radica en que el agente kantiano desatiende la integridad en la medida en que privilegia el punto de vista racional. Sin embargo, creo que se puede ir un poco más lejos y articular una crítica un poco más fuerte. No solo pienso que la integridad entra en tensión con la concepción del yo dividido, sino que, además, creo que esta concepción es estructuralmente deficiente para darle sentido a la idea misma de integridad personal.

La tensión planteada entre la integridad y el modelo kantiano de la agencia tiene el supuesto de que el carácter y los proyectos fundamentales de una persona se ubican en la parte fenoménica del yo kantiano. ¡Dónde más podrían estar! En virtud de este modelo, todo lo que no hace parte del yo racional será un *residuo* fenoménico. Y es verdad que el carácter incluye muchos elementos, entre los cuales encontramos deseos, intenciones, planes, proyectos fundamentales, etc. Muchos de estos elementos bien podrían ubicarse en el aspecto fenoménico del yo; y en términos kantianos, muchos de ellos bien podrían caracterizarse como inclinaciones[61]. No obstante, y esta es la idea que busco precisar, elementos específicos como los deseos categóricos o los proyectos fundamentales no pueden ser tratados sin más como inclinaciones. Los deseos y proyectos juegan un papel más fundamental que una simple inclinación, pues, como ya he indicado, la satisfacción de las

[61] "Kant define la 'inclinación' como un deseo que se ha convertido en hábito (*MC* 15). En tanto que 'deseo', la inclinación se origina en los sentimientos que experimentamos. […] [D]ebido a que las inclinaciones se originan en nuestra constitución natural sensible, no tienen por qué coincidir naturalmente con las exigencias del deber, las cuales se originan en la razón pura" (Rivera, 2014, p. 37).

inclinaciones está condicionada a que estemos vivos; en cambio, los deseos categóricos y los proyectos fundamental explican por qué seguimos vivos (§4.2.1.b). Si se mantiene la distinción entre razón-inclinación, simplemente la idea de integridad no tendría sentido. Por esta razón, pienso, el problema, en una versión fuerte, no radica tanto en que la agencia kantiana amenace la integridad personal, sino que, literalmente, el agente kantiano es ciego a su integridad.

El argumento entonces muestra que un agente genuinamente preocupado por su integridad no es un yo racional, pero tampoco es un yo que sucumbe ante las inclinaciones. No es un yo racional porque este carece de carácter individual; y tampoco es un yo empírico, porque los deseos categóricos o los proyectos no equivalen a inclinaciones. La naturaleza de los estas categorías de la integridad no puede ser bien capturadas por el modelo kantiano de la agencia. Incluso si se acepta que los deseos categóricos o los proyectos fundamentales pertenecen a la parte fenoménica, habría que inventar una nueva categoría, más allá de las inclinaciones, en dónde ubicarlos. Y aun así, digo, sigue siendo claro que la integridad se vería amenazada por las exigencias del "buen orden imparcial del mundo de los agentes racionales". Dicho esto, resulta que la concepción del yo dividido no puede darle sentido a lo que significa estar genuinamente preocupado por la propia integridad, pues este modelo de agencia no hace inteligible la idea de los deseos categóricos o de los proyectos fundamentales de una persona.

En conclusión, el problema de la integridad no es secundario o menor, sino que representa un serio cuestionamiento al modelo kantiano de la agencia. En una versión débil de mi crítica, la concepción del yo dividido no es compatible con el concepto de integridad; y en una versión más fuerte, esta concepción simplemente hace ininteligible la preocupación de una persona por su integridad. El reto de la posición kantiana, entonces, es poder darle sentido a la noción de integridad.

III. ¿Supera el modelo kantiano el problema de la integridad?

5. Christine Korsgaard: integridad como auto-constitución

5.1 Agencia e integridad
5.2 Modelo constitucional de la agencia
5.2.1 modelo constitucional de la agencia
5.2.2 El modelo constitucional y el agente kantiano
5.2.3 Identidades prácticas
a. Contingencia
b. Identidad práctica humana
c. Identidad moral
5.2.4 Razones públicas y razones privadas
a. La naturaleza pública de las razones
b. Ley moral
c. Integridad y unidad
5.3 Integridad como auto-constitución

El problema de la integridad personal fue planteado por B. Williams como una crítica específica en contra del utilitarismo. En el capítulo anterior he querido ampliar su crítica para mostrar que el rigorismo moral kantiano también se ve afectado por este problema. Básicamente, se trata de un problema que se le presenta a toda concepción moral que imponga un punto de vista imparcial e impersonal a costa de un punto de vista personal. El rigorismo kantiano sería un ejemplo de este tipo de concepción moral. La imparcialidad de los deberes morales y la exigencia de su cumplimiento incondicional conducen a que los agentes, en ocasiones, tengan que renunciar a su carácter o a sus proyectos fundamentales, y en esa medida el rigorismo representa una amenaza a la integridad de la persona.

Barbara Herman, en su artículo "Integrity and Impartiality" (1983), plantea una estrategia para responder al problema de la integridad personal. Según esta autora, el problema tiene repercusiones importantes para nuestra comprensión de la vida moral, pues nos deja en una situación de "vulnerabilidad moral": para sostener y realizar nuestros proyectos

fundamentales tenemos que estar dispuestos a realizar algo que la moralidad imparcial condena (*cf.* Herman, 1983, p. 241). Lo que se desprende de la preocupación por la integridad personal, dice Herman, es que el hecho de saber que un determinado curso de acción está mal moralmente no necesariamente nos da una razón suficiente para inhibirnos de seguirlo. Según esta autora, "si la moralidad impacial puede interferir con el propósito de los proyectos fundamentales de una persona, habrá casos donde un agente podría no tener razón para actuar como la moral requiere, ya que las únicas razones para actuar son aquellas que lo dirigen a perseguir su proyecto fundamental" (*ibid.*).

La estrategia de Herman para responder al problema de la integridad consiste en señalar que la moralidad (imparcial) no es algo ajeno al carácter de los agentes, pues "el compromiso [*attachment*] con la moralidad imparcial puede ser en sí mismo un proyecto que le da sentido a la vida […] vivir moralmente puede ser una parte constitutiva del carácter" (Herman, 1983, p. 242). No es necesario entonces que la moralidad imparcial entre en conflicto ni con el carácter ni con los proyectos fundamentales de una persona. En mi interpretación, esta estrategia no es suficiente para responder al problema de la integridad. Es verdad que la moralidad puede llegar a formar parte del carácter, es verdad también que la moralidad puede ser uno de los proyectos fundamentales de alguien, pero esto no es suficiente para que no se presente una posible amenaza a la integridad. El origen del problema radica en que el rigorismo moral sostiene la prioridad normativa que tiene el deber, es decir, la moralidad debe imponerse por encima de cualquier otro criterio práctico. Incluso si la moralidad es uno de los proyectos fundamentales de alguien, la amenaza contra su integridad persiste, pues lo que la imagen kantiana de la agencia pretende es que la moral sea un valor incondicional; y es por eso que puede entrar en conflicto con el carácter y con otros proyectos fundamentales. La exigencia del rigorismo kantiano es que el agente esté dispuesto a dejar de lado sus proyectos, si estos implican acciones que vayan en contra del deber.

De manera que una respuesta al problema de la integridad no solo tiene que mostrar cómo la moralidad se integra al carácter y a los proyectos; también tiene que justificar por qué la moralidad tiene o debe tener una prioridad o primacía normativa por encima de otros proyectos personales. Y de hecho, Herman de alguna manera es consciente de la necesidad de ofrecer una justificación de la prioridad normativa de la moralidad. Esta autora piensa

que la moralidad kantiana, a través del imperativo categórico, funciona como un proyecto fundamental que sirve como ideal regulativo para el resto de proyectos que alguien pueda tener. Para ella, y dicho de manera general, el imperativo categórico no sirve para evaluar máximas de acción (ver §2.2), sino que tiene la función de establecer condiciones limitantes para la acción. De esta manera, una acción o un proyecto serían permisibles moralmente, siempre que no violen el límite moral impuesto por el imperativo categórico. Esto permite que el agente tenga un compromiso con la moralidad imparcial, sin que necesariamente todos los proyectos fundamentales tengan que ser abandonados, pues solo habría que rechazar aquellos que resulten impermisibles moralmente. No es necesario entonces que alguien actúe únicamente según valores imparciales, ya que él o ella puede estar comprometido con proyectos muy personales que no necesariamente están en contra del ideal regulativo moral.

Según mi lectura, el punto más fuerte de la propuesta de Herman es la idea de que los proyectos fundamentales no son normativamente incondicionales (Herman, 1983, p. 243), pues todos ellos están determinados por las condiciones limitantes que establece la moralidad.

> El argumento kantiano es que en el límite, donde el conflicto con la moralidad es serio e inevitable, la moralidad debe ganar. La 'victoria' de la moralidad no disminuye el valor del proyecto (a menos que el proyecto en sí mismo sea inmoral), y no constituye un ataque a la integridad. Más bien se trata de una indicación de que en este mundo, o en estas circunstancias, no hay una forma permisible de continuar con el proyecto. (Herman, 1983, p. 243)

La permisibilidad o impermisibilidad de un proyecto está condicionada entonces por la moralidad. La moralidad ha de "formar y limitar otros proyectos de manera que sean compatibles con un sentido ideal de cómo una persona debería vivir" (Herman, 1983, p. 244). De esta manera, la solución de Herman devuelve el privilegio normativo al deber moral. En otras palabras, mientras que Williams se esfuerza por mostrar que la integridad en ocasiones puede ser incompatible con la moralidad, Herman responde que la integridad ha de estar limitada por la moralidad. En una palabra, la integridad sería un valor condicionado por la moralidad.

La posición de Herman es una estrategia compatibilista, pues intenta conciliar la integridad

y la moralidad: la integridad está limitada, eso sí, por un marco moral kantiano. Tener un carácter no puede ser pensado al margen del compromiso con la moralidad. Por eso, para Herman, estar condicionado por la moralidad, establecer los límites de lo permisible y lo impermisible "es, en nuestro sentido ordinario de hablar, un asunto de integridad" (Herman, 1983, p. 244).

> "[L]a moralidad kantiana puede ser (y debe ser tomada como) definitoria de un sentido del yo, y que al tener un *carácter moral*, una persona no tendrá que abandonar algo en el sentido de la integridad que se ubique más allá de lo que la moralidad imparcial permita". (Herman, 1983, p. 244; énfasis añadido)

El argumento de Herman sin duda responde a la crítica de Williams. Sin embargo, la fuerza del argumento depende de qué signifique que la moralidad sea un ideal regulativo. Esta autora se ha esforzado por explicar este punto, especialmente en "Moral Deliberation and the Derivation of Duties" (1993a). Que la moralidad kantiana sea un ideal regulativo significa que es un criterio negativo para establecer la permisibilidad de una acción[62]. Sin embargo, incluso si esta interpretación de la filosofía moral kantiana es correcta, Herman no explica por qué el ideal regulativo es necesariamente vinculante para una persona. Herman asume que el agente está ya comprometido incondicionalmente con la moralidad[63]. Este supuesto la lleva a asociar (i) la idea de que la moralidad es un ideal regulativo con (ii) la idea de que la persona está incondicionalmente comprometida con la moralidad. Pero esta es una asociación errónea: que algo sea un ideal regulativo es una razón, tal vez, para querer cumplirlo, pero no es una razón para estar comprometido incondicionalmente a cumplirlo. Herman tiene entonces el reto de explicar en qué sentido una persona está incondicionalmente comprometida con la moralidad regulativa. Y una vez sea explicado el sentido de este compromiso, será posible entender en qué sentido el resto de proyectos están condicionados a este ideal moral. Sin embargo, tal y como entiendo la posición de Herman,

---

[62] "El papel del motivo kantiano del deber es como tal una condición limitante: expresa el compromiso de un agente de que él no actuará (bajo cualquier motivo, o cualquier fin), a menos que su acción sea moralmente permitida" (Herman, 1983, p. 236).

[63] "[E]l compromiso con la moralidad se supone que *es* incondicional. El agente moral es aquel que tiene una concepción de sí mismo como alguien que no persigue sus proyectos de formas moralmente impermisibles" (Herman 1983, p. 243s; énfasis en el original).

ella no logra cerrar esta brecha entre postular la moralidad como un ideal regulativo y el compromiso efectivo e incondicional que una persona podría tener con la moralidad.

A diferencia de Herman, Christine Korsgaard sí asume el reto de explicar la incondicionalidad de la moralidad y, por lo tanto, ofrece una de las respuestas más elaboradas e interesantes al problema de la integridad personal. Su propuesta, que comparte la idea de Herman de que la moralidad es un valor incondicional en la vida práctica humana, se desarrolla sobre todo en *Self-constitution. Agency, Identity and Integrity* (2009). En este libro, la autora no plantea el problema en términos del rigorismo moral kantiano, tal y como lo hago en esta investigación, y, sin embargo, definitivamente ella sí se compromete con la idea de una primacía normativa de la moralidad[64], que es la idea que está a la base del rigorismo moral kantiano. Desarrollaré este concepto más adelante (§6.2). Por ahora, basta mencionar que la primacía normativa de la moralidad es la tesis que sostiene que la moralidad tiene mayor peso normativo que cualquier otro criterio práctico. Esta tesis es el corazón del rigorismo moral kantiano.

Ahora bien, la respuesta de Korsgaard al problema de la integridad descansa en la articulación de dos ideas. Primero, ella se compromete con la idea de que para ser agente es necesario ser un "todo unificado" o un ser integrado. En este sentido, para esta autora, la integridad debe entenderse en términos de unidad. Esta interpretación está justificada a la luz de su modelo agencial, el modelo constitucional. Segundo, según Korsgaard, la integración o unificación de la agencia solo es posible a través del compromiso del agente con principios morales universales.[65] Con esta idea, Korsgaard defiende entonces que las exigencias de la moralidad no fragmentan al individuo; al contrario, hacen posible su unidad

[64] Por citar uno de muchos pasajes: "Nosotros solo podemos vincular [*attach*] el 'Yo quiero' [*I will*] a nuestras decisiones si queremos nuestras máximas como leyes universales. El imperativo categórico es el principio constitutivo de la acción, según Kant, porque la conformidad con él es constitutiva del ejercicio de la *voluntad*, de la determinación de una persona *por sí misma* en oposición a su determinación por algo dentro [o fuera] de ella" (Korsgaard, 2009, p. 76; énfasis en el original).

[65] "Voy a argumentar que los principios de la razón práctica son principios por medio de los cuales nos constituimos como agentes unificados. Y argumentaré que eso explica su normatividad. Los principios de la razón práctica nos obligan [*bind*] porque, al tener que actuar, debemos constituirnos a nosotros mismos como agentes unificados" (Korsgaard, 2009, p. 25).

o integridad. La consecuencia de estas dos ideas es que no es posible ser un agente a menos que se tengan compromisos con principios morales. Como lo interpreto, esta conclusión es una defensa de la tesis de la primacía normativa de la moralidad: la moralidad es tan importante para un agente que sin ella se desintegraría y, literalmente, dejaría de ser agente. Dicho de otra manera, la verdadera amenaza a la integridad consiste en abandonar la moralidad.

A diferencia de la posición de Williams, para Korsgaard la moralidad y la integridad personal no pueden entrar en tensión porque la moralidad es una condición de la integridad. En este capítulo me interesa examinar en detalle esta respuesta. En el capítulo siguiente, espero hacer una crítica a la propuesta de Korsgaard y a la vez desarrollar en un sentido más preciso la forma en la que la integridad representa un problema para la teoría kantiana.

5.1 Agencia e integridad

¿Qué significa ser agente? En un sentido muy básico e intuitivo, ser agente significa actuar. Un agente es quien actúa en el mundo. Actuar no es lo mismo que ser movido como un objeto o ser un simple medio en una cadena causal. Ser agente consiste en ser un tipo especial de causa. De esta intuición es la que parte Korsgaard para desarrollar su teoría de la agencia. Y como vemos, se trata de una intuición muy poderosa: la condición de ser agente es indesligable de la acción. Para entender la agencia habría que entender la naturaleza de la acción.

En la tradición filosófica se ha asumido que la agencia es una condición ya dada, que ser-agente es el punto de partida. Asumiendo esto, la pregunta tradicional consiste en explicar cómo actúa un agente que ya está constituido. Korsgaard, por su parte, considera que ser-agente es algo que no puede darse por sentado. La agencia es algo que se construye, es una tarea por cumplir, y por tanto, antes de preguntarse cómo actúa un agente, habría que preguntarse cómo se llega a ser-agente.

Korsgaard sostiene que llegar a ser-agente significa ser autor de las propias acciones. Esto es apenas obvio, dada la relación entre agencia y acción. Ser autor significa ser aquel a quien se

le atribuyen los actos y los movimientos que se llevan a cabo (Korsgaard, 2009, p. 18). La acción de un autor no puede ser el producto de alguna fuerza que opera fuera o incluso dentro de alguien. Las fuerzas externas o internas son ajenas a la agencia y, desde esta perspectiva, son vistas como las causa de los eventos que le suceden al agente. Un espasmo, un calambre, un desmayo, son eventos causados por alguna parte del yo (un músculo, un nervio, una deficiencia de azúcar); pero no son fenómenos que correspondan al autor de las acciones. Por el contrario, la acción de un autor debe ser la expresión del 'yo como un todo':

> Para que un movimiento cuente como mi acción, debe ser expresivo de mí mismo en la forma en que una acción debe serlo, debe ser el resultado de mi entera naturaleza operando como un *todo integrado* [*integrated whole*]. (Korsgaard, 2009, p. 19; énfasis añadido)

Para los propósitos de esta investigación, la idea más interesante de la propuesta de Korsgaard es la afirmación según la cual no es posible llegar a ser autor o agente de una acción si el agente no está unificado o integrado. En otras palabras, la integridad o unidad es uno de los rasgos definitorios de la agencia: si alguien no es un todo integrado, no podría tampoco ser autor de sus acciones, o lo que es lo mismo, la acción solo es posible en la medida en que su autor esté integrado.

En contra de la visión tradicional de la agencia, Korsgaard considera que no existe previamente tal yo integrado. La integración de la agencia es una tarea por realizarse. Por ello es que no se puede pensar la acción como el resultado de un autor o de un agente cuya integridad ya está dada o garantizada de antemano. Aunque, según Korsgaard, la acción necesita de un agente unificado, esta unificación se constituye (construye) a través de las acciones y de las decisiones. Esto quiere decir que llegar a ser agente comporta un proceso de integración y que este proceso se instancia en cada elección y en cada una de las acciones que el agente realiza[66] (Korsgaard, 2009, p. 19).

---

[66] Esta formulación, tal y como Korsgaard lo reconoce, conlleva a la *paradoja de la auto-constitución*: "¿Cómo puedes auto-constituirte, crearte a ti mismo, a menos que ya estés ahí?" (Korsgaard, 2009, p. 20; énfasis añadido). Si se renuncia a la idea tradicional de que el agente es una entidad ya dada, y en cambio se piensa que la agencia es una tarea o una actividad, entonces surge el problema de

Para Korsgaard, como vemos, existe una conexión intrínseca entre la agencia y la integridad. Si ser agente consiste en actuar, si actuar supone ser autor de la acción, y si para ser autor es necesario ser un todo integrado, entonces, por transitividad, ser agente es ser un todo integrado [*whole integrated*].

Es natural esperar que esta tesis tenga repercusiones para el problema de la integridad. Como quedó claro en el capítulo anterior (§4.2), Williams interpreta la integridad en términos de carácter. Y en este sentido, una amenaza a la integridad es una amenaza al carácter y los proyectos de la persona. En contraste, Korsgaard interpreta la integridad en términos de unidad; un sentido muy cercano a la etimología de la palabra integridad: *completo*, *entero*, *unificado* (*integer*).

> Etimológicamente, la integridad es unidad [*oneness*], la integración es lo que hace que algo sea uno. Ser algo, una cosa, una unidad, una entidad; ser algo en absoluto: en el sentido metafísico, eso es lo que significa tener integridad. (Korsgaard, 1996, p. 102)

La interpretación del concepto de integridad de Korsgaard, evidentemente, difiere de la interpretación que hace Williams (§4.2). De hecho, se trata en cada caso de conceptos tan distantes que podría parecer que no hay ningún vínculo entre ellos. ¿Qué relación podría existir entre estas dos interpretaciones? ¿Qué relevancia tiene la interpretación de Korsgaard frente al problema de la integridad planteado por Williams? Parte de mi interés es justamente construir el diálogo que podría establecerse entre los dos autores. El objetivo de este capítulo es entonces explicar la interpretación de Korsgaard del concepto de integridad (§5.2), y posteriormente mostrar en qué medida esta interpretación podría responder al problema de la integridad planteado por Williams (§5.3). Como veremos, tanto Williams como Korsgaard atienden a los mismos interrogantes: ¿qué es la integridad y por qué es tan importante en nuestra comprensión de la agencia humana?

---

explicar *quién* es el que realiza la actividad sin que haya un agente previo. En §5.2.2 quedará explícita la solución a esta paradoja.

5.2 Modelo constitucional de la agencia

En §4.3 quedó claro que el problema de la integridad personal se encuentra anclado al modelo de la concepción del yo dividido. Y como concluí, la idea de integridad no es compatible con este modelo de la agencia. Por esta razón, si Korsgaard mantiene la concepción del yo dividido, no podría darle sentido a la noción de integridad. ¿Cómo es posible darle sentido a la integridad dentro del paradigma de la filosofía kantiana? Según mi interpretación, Korsgaard responde a esta pregunta en dos momentos. El primero consiste en reinterpretar la noción de integridad. Para Korsgaard, repito, la integridad ha de ser entendida en términos de unidad. Un yo integrado es un yo unificado. Esta interpretación del concepto, sin embargo, es abiertamente contradictoria con el modelo de la agencia del yo dividido. Si el modelo agencial kantiano está estructuralmente comprometido con un yo dividido, entonces la pretensión de un yo unificado resulta casi impensable. Si la condición del agente kantiano es estar dividido, entonces la pretendida integridad que busca Korsgaard resultaría contraria a esta condición estructural. Por esta razón, el segundo momento de su respuesta consiste en rechazar la concepción del yo dividido y, en cambio, caracteriza un nuevo modelo de la agencia kantiana. En este sentido, ella propone interpretar la agencia kantiana a través del *modelo constitucional de la agencia*.

5.2.1 Modelo constitucional de la agencia

Korsgaard considera que la concepción del yo dividido surge por un compromiso teórico con lo que ella denomina el *modelo de combate*. La concepción del yo dividido tiene como rasgos distintivos la distinción entre un yo pasivo y un yo activo, y la prioridad normativa del yo activo sobre el pasivo (§3.4). El modelo de combate sería una forma específica de esta concepción que describe el conflicto que existe específicamente entre la razón y la inclinación: en la filosofía kantiana —según el modelo de combate— existe una oposición constante entre la razón y las inclinaciones, y el agente experimenta un conflicto constante en el que las partes luchan por imponerse una sobre la otra. Una caracterización muy elocuente de este conflicto la encontramos en las siguientes palabras de Hume:

> Nada es más corriente en la filosofía, e incluso en la vida cotidiana, que el que, al hablar del combate entre pasión y razón, se otorgue ventaja a esta última, afirmando que los hombres son virtuosos únicamente en cuanto que se conforman a los dictados de la razón. Toda criatura racional, se dice, está obligada a regular sus acciones mediante la razón. Y si algún otro motivo o principio desafía la dirección de la conducta de esa persona, ésta tendrá que oponerse a ello hasta someterlo por completo, al menos hasta conformarlo con aquel principio superior. (Hume, 1976, p. 614)

A primera vista, la filosofía práctica kantiana sería el ejemplo perfecto del modelo de combate. No son pocos los lugares en los que se puede encontrar evidencia textual al respecto:

> En cuanto fuentes de necesidades, las inclinaciones mismas distan tanto de albergar un valor absoluto para desearlas por ellas mismas, que más bien ha de suponer el deseo universal de cualquier ser racional el estar totalmente libre de ellas. (F 4:428)

El modelo de combate refleja bien este conflicto entre la razón que es imparcial y objetiva, y las inclinaciones y deseos que son caprichosos y arbitrarios. La razón es sinónimo de autodominio, en cambio "las inclinaciones son fenómenos no cognitivos que están fuera de nuestro control" (Oakley, 1990, p. 443). Justamente este carácter aleatorio y caprichoso de las inclinaciones las convierte en algo que necesita ser controlado, y es aquí donde se justifica la idea de un combate entre una razón que controla y unas inclinaciones que han de ser controladas. El modelo de combate incluso encuentra soporte en la metafísica kantiana: las inclinaciones hacen parte de ese mundo causal y fenoménico y por tanto son cosas que padecemos (*pathos*) los seres humanos. Al dejarnos llevar por ellas nos convertimos en objetos del mundo causal y perdemos nuestra condición de agentes libres (§3.2). La razón práctica, por su parte, nos libera de esta causalidad fenoménica (libertad negativa) y a través de ella podemos auto-legislarnos y, por tanto, ser agentes libres (libertad positiva).

La interpretación más común y tradicional de la filosofía moral kantiana asume que el modelo de combate es la descripción adecuada del agente kantiano. Sin embargo, Korsgaard cree que este modelo representa una lectura incorrecta de la agencia kantiana. Según esta

autora, en dicho modelo la razón y las inclinaciones son vistas como fuerzas opuestas que luchan por imponerse una sobre la otra. Y cuando una de ellas gana, entonces se procede a la acción. El problema de esta versión es que el agente desaparece del escenario, o más bien, la agencia queda reducida a un juego de fuerzas, pues la acción sería simplemente el resultado de la lucha entre la razón y las inclinaciones (*cf.* Korsgaard, 2009, p. 134).

Si el modelo de combate es inadecuado para entender la agencia kantiana, cuál es, entonces, el modelo que realmente da cuenta de esta forma de entender la agencia. Para responder a esta pregunta Korsgaard propone el *modelo constitucional de la agencia*. El punto de partida de este modelo es el hecho de que somos agentes auto-conscientes:

> Tanto los seres humanos como los otros animales actúan, pero las acciones humanas pueden ser correctas o incorrectas moralmente, mientras que las acciones de los otros animales no. Esto debe ser porque hay algo distintivo en la naturaleza de la acción humana. […] El nombre que le doy a esta característica distintiva es la tradicional —racionalidad. Como lo entiendo, la razón es un poder que tenemos en virtud de cierto tipo de auto-consciencia —conciencia de los fundamentos de nuestras propias creencias y acciones. Esta forma de auto-consciencia nos da la capacidad para controlar y dirigir nuestras creencias y acciones que los otros animales no tienen, y nos hace activos en una forma que ellos no lo son. Pero también nos genera un problema que los otros animales no enfrentan —el problema de decidir qué cuenta como una razón para creer o para actuar. Para ponerlo de otra manera, esta forma de auto-consciencia hace necesario tomar el control de nuestras creencias y acciones, pero debemos entonces esforzarnos en cómo lo conseguimos: debemos encontrar principios normativos, leyes, que gobiernen lo que creemos y hacemos. (Korsgaard, 2009, p. xi; énfasis en el original)

El hecho de que seamos autoconscientes significa que podemos tomar distancia de nuestras creencias, deseos, incentivos o fuerzas. En este sentido, la autoconsciencia trae como consecuencia una fragmentación en el agente o, en términos platónicos, la idea de que el

alma tiene partes[67]. El agente autoconsciente se ve entonces enfrentado a preguntarse qué es lo que va a hacer con respecto a estos estados mentales o a estos fenómenos corporales que le incitan a moverse en una determinada dirección. Él toma el control de estos fenómenos solo en la medida en que adopta una distancia reflexica con respecto a ellos. Como se ve, la autoconsciencia supone una división: una parte autoconsciente, y otra que es objeto de esta autoconsciencia.

Korsgaard cree que la división del agente, así entendida, no tiene por qué ser concebida según el modelo de combate, sino según el modelo constitucional. Este modelo está inspirado en la teoría del alma que ofrece Platón en la *República*. En líneas generales, Platón traza una simetría entre el alma humana y la *polis* griega. Un alma con buen orden, es decir, virtuosa, se asemejaría a una *polis* justa. Para Platón, el estado justo ha de estar divido en clases: una clase gobernante que crea las leyes, una clase de soldados cuya función es hacer cumplir las leyes y proteger la ciudad, y finalmente, una clase de artesanos cuya función es cubrir las necesidades de todos los ciudadanos. "Las virtudes de una *polis* ideal se identifican con ciertas propiedades y relaciones entre estas partes" (Korsgaard, 1999, p. 5). Platón considera que el ideal de una ciudad es alcanzar la justicia. Y para él, la justicia consiste en que cada parte que constituye el estado cumpla adecuadamente su función. "La justicia ha de consistir en hacer lo que corresponde a cada uno, del modo adecuado" (*Rep.* 433a-b).

La función de cada una de las partes del estado estaría determinada por su función en la constitución misma del estado. Los legisladores, los soldados o los artesanos cumplen su función en la medida en que contribuyan a la formación de la *polis*. Un artesano, por ejemplo, suple las necesidades de la *polis* porque está pensando en la constitución del estado como un todo, y no tanto porque esté ejercitando una técnica que tenga valor por sí misma. El

[67] "La autoconciencia introduce otra diferencia más inmediata en la vida de los animales humanos. Introduce lo que, siguiendo a Platón, llamaré las partes del alma […]. El alma de los animales no-humanos no tiene partes, ya que lo que cuenta para un animal no-humano como un incentivo es establecido por sus instintos, y sus instintos inmediatamente le dicen cómo tratar con el incentivo: en el alma animal el trabajo de incentivo y el trabajo del instinto no son fenomenológicamente diferentes. Así, en un animal no-humano, el sistema psíquico completo está cerrado y rigurosamente tejido. Y eso significa que hay un sentido en el que la auto-conciencia *produce* las partes del alma" (Korsgaard, 2009, p. 119; énfasis en el original).

legislador promulga leyes, no por promulgarlas sin más, sino porque está pensando en el bien de la *polis* en general. Igualmente, el uso de la fuerza por parte de los soldados siempre está al servicio de la protección del estado, no se trata de una fuerza ciega o arbitraria. A partir de esta interpretación, Korsgaard concluye que lo que legitima la *polis*, en últimas, es la idea de *constitución*. Pensar el estado como una totalidad constituida es lo que unifica las distintas partes y además es lo que legitima la función de cada una de ellas dentro de la *polis*. "La constitución hace posible para un grupo de ciudadanos —que sin la constitución, serían un mero amasijo de personas— que funcione como un agente colectivo individual" (Korsgaard, 2009, p. 142).

Sería ingenuo pensar que siempre hay una armonía entre las partes que componen la *polis*, de ahí que sea necesaria cierta virtud política: la *sophrosyne* (*Rep.* 430e-432b). Esta virtud, en el Platón de la *República*, se entiende como autocontrol o autodominio. Y, advierte Korsgaard, el autodominio no debe ser entendido como un modo de represión o constreñimiento forzado, pues esta virtud supone un acuerdo, una creencia compartida por todas las clases de la ciudad, acerca de cuál es la clase que debe dominar (*Rep*.431d–e; 442c–d). De esta manera, en la ciudad-estado de Platón todas las partes involucradas aceptan el dominio de la clase gobernante, cuyos miembros, en nombre de la constitución del estado, actúan, o al menos intentan actuar racionalmente.

A partir de esta analogía con la idea de la *polis* concebida por Platón, resulta claro que el modelo de combate sería la descripción de una *polis* en guerra civil. En estas condiciones las distintas clases están en combate, de manera que cuando una de ellas sale victoriosa, el tipo de relación que se establecería entre ellas sería tiránica e impositiva. Por su parte, el modelo constitucional supone un acuerdo entre las clases que conforman el estado, de manera que cuando surge un conflicto, es trabajo del gobernate, con ayuda de los soldados, aplacar la rebelión. En el modelo constitucional, la autoridad de la razón no se describe como una guerra interna o como una tiranía, sino como un ejercicio legítimo de poder en nombre de la constitución (i.e. integridad) del estado.

Cuando se traduce esta teoría del estado a una teoría de la agencia, se da por sentado, como

ya he mencionado, que el alma está dividida o que tiene partes. Bajo el modelo de combate, la razón sería una tirana que ejerce su poder en contra de las partes inferiores del alma. La legitimidad de la razón solo dependería de la potencia de la fuerza que tenga para aplacar la rebeldía de las pasiones o las inclinaciones. En el modelo de combate, lo que tenemos es la descripción psicológica del "pecador miserable reformado" [*Reformed Miserable Sinner*] cuyos deseos han de ser reprimidos en nombre de la razón (Korsgaard, 2009, p. 147). En el modelo constitucional, en cambio, no hay una guerra interna; lo que se busca, más bien, es la unidad —o la armonía— de las partes: una persona constituida tendría todas sus partes unificadas o integradas gracias al ejercicio de una razón que provee orden y forma a las distintas partes que la constituyen. Es muy importante aclarar que en el modelo constitucional el agente no elige la razón porque esta deba ganar el conflicto, o porque la razón sea moralmente preferible al instinto, sino porque el agente necesita estar constituido, y la idea de constitución implica, para él, que su razón debe gobernar sobre el instinto. El agente está dispuesto hacia la razón porque esta es la que hace posible que él pueda auto-constituirse. "Si el agente se conforma al dictado de la razón, no es porque se identifique con la razón, sino más bien porque se identifica con su constitución, y esta dice que la razón debería gobernar" (Korsgaard, 2009, p. 135).

La pregunta que surge inmediatamente es por qué un agente tendría que preocuparse por su constitución, por qué esta es tan importante. A la luz de la analogía política, un estado necesita estar constituido para ser considerado un estado como tal. Cuando decimos que Atenas condenó a Sócrates, no pensamos que Atenas sea un agente individual, sino que se trata de una entidad que, aunque esté compuesta por varios individuos, actúa como un todo. La constitución del estado es lo que diferencia *una polis* de una mera agregación de individuos, por lo tanto, un estado necesita estar constituido para ser un estado. En este mismo sentido, y continuando con la analogía, la constitución del agente es lo que permite que él sea un agente.

Si ser agente es ser un todo integrado, entonces sus acciones son muy distintas a un mero comportamiento irreflexivo o a un movimiento físico (p.e. un reflejo o un proceso fisiológico). La diferencia radica entonces en que las acciones tienen siempre a un autor (Korsgaard, 1999,

p. 3; 2009, p. 18), mientras que los movimientos automáticos o los reflejos carecen de reflexión o son respuestas causales que corresponden a las partes de la persona, no a ella como un todo.

Todo el argumento de Korsgaard va encaminado a demostrar que, a la luz del modelo constitucional, la constitución es la condición de posibilidad de la agencia. Y solo en virtud de que el agente está constituido es que se le puede atribuir integridad, sin constitución no habría totalidad y, por lo tanto, no habría agente unificado. 'Autor' es el nombre para designar esta totalidad integrada o esta unidad del agente. Además, siguiendo la analogía con el estado justo de Platón, resulta que la constitución y la unidad solo son posibles a través del ejercicio de la razón, pues es esta facultad la única capaz de establecer los procedimientos y las reglas para integrar las distintas partes del alma (Korsgaard, 1999, p. 10).

Con este análisis se entiende entonces la relación tan estrecha entre la agencia y la integridad: la integridad (unidad) es condición de posibilidad de la agencia. Es claro que esta conclusión, de cuño trascendental, no tiene mucho sentido dentro de la concepción del yo dividido; pero a la luz del modelo constitucional que propone Korsgaard, resulta comprensible que la integridad sea indesligable de nuestra idea de ser agentes[68].

5.2.2 El modelo constitucional y el agente kantiano

Según Korsgaard, el modelo constitucional, ya presente en Platón, también puede encontrarse en la filosofía práctica de Kant. No es evidente, sin embargo, la manera en que

[68] El modelo constitucional no solo ofrece una teoría de la agencia, también es una propuesta para establecer una normatividad de la acción. Según Korsgaard, este modelo también ofrece un criterio para las acciones buenas, "un criterio que nos dice cuáles acciones son verdaderamente más propias de la persona, y por tanto, qué acciones son más verdaderas" (Korsgaard, 1999, p. 3). Según Korsgaard, las acciones que son "más propias" de una persona son aquellas que "más unifican completamente" y por tanto "más constituyen completamente" al agente como su autor. En este sentido, Korsgaard apuesta por que su modelo constitucional configure la unidad volitiva del agente, de manera que sus acciones puedan ser atribuidas a él como un todo. Esta idea es muy interesante, en especial porque es un intento por establecer la relación entre el modelo de agencia y la unidad volitiva. Más adelante mostraré las razones por las que, sin embargo, considero que esta propuesta es equivocada.

la filosofía práctica kantiana puede ser interpretada a la luz del modelo constitucional.

Korsgaard rechaza una interpretación empirista de la filosofía práctica de Kant, según la cual las inclinaciones son atracciones (mecánicas) hacia objetos del mundo. En vez de eso, ella propone que uno de los conceptos centrales para entender la motivación es el incentivo o impulso [*Triebfeder*[69]]. Y distingue el incentivo de la inclinación:

> [C]uando eres consciente de la operación de un incentivo natural sobre ti –consciente de que algo te atrae, o te tienta a la acción– te describes a tu mismo como 'inclinado' a actuar en la forma en que el incentivo tienta. Así que eso es lo que una inclinación es: es tu consciencia de que estas naturalmente atraído a actuar de cierta manera. (Korsgaard, 1998, p. 50s)

Los incentivos son fenómenos causales y, sin embargo, por sí mismos no causan la acción. Hace falta además la consciencia de este incentivo (inclinación) y también hace falta que el agente asienta o eija [*choose*] a esta inclinación (*ibid.* p. 51). Los incentivos trabajan en conjunto con los principios que "determinan la respuesta de los seres conscientes a estos incentivos" (*ibid.* p. 50). En esta medida, el incentivo no es suficiente para que se constituya como un motivo para una acción, hacen falta principios que guíen la satisfacción o no de este incentivo. "Un motivo, podría decir, es un incentivo operando bajo un cierto principio, o visto desde la perspectiva ofrecida por ese principio" (*ibid.*).

Aclaro que para Korsgaard la consciencia —que no debe ser confundida con auto-consciencia— es un fenómeno que también es propio de los animales no humanos. Y los incentivos también están presentes en ellos, solo que su respuesta es automática: si sienten hambre, el instinto los lleva a cazar; si sienten miedo, el instinto les hace huir. En cambio, los seres humanos, en la medida en que somos *auto*-conscientes, podemos tomar distancia reflexiva con respecto a los incentivos, es decir, podemos ser conscientes de los efectos de esos incentivos sobre nosotros. Y cuando somos conscientes de este modo —conscientes de que algo nos atrae o nos resulta tentador—, entonces nos consideramos "inclinados" a actuar

[69] Para una ampliación del concepto de incentivo, ver Korsgaard (1996, p. 242ss).

para procurar lo que el incentivo ofrece (*cf.* Korsgaard, 1998, p. 51). "Esto es lo que es la inclinación: es la consciencia de que tú estás naturalmente atraído a actuar de cierta manera" (*ibid.*).

En este espacio creado por la auto-consciencia es donde surge la pregunta por las razones para actuar: cuando te preguntas si actuarás de acuerdo con la inclinación causada por el incentivo, realmente estás preguntando si tu inclinación es una *razón* para actuar. La autoconsciencia de los propios estados mentales, la distancia reflexiva que tomamos respecto de ellos, es la fuente de la libertad de la voluntad [*free will*]: en la medida en que somos conscientes de que estamos inclinados a actuar según cierto incentivo, podemos decidir si actuar o no (§3.4).

Esta nueva imagen de la psicología kantiana, continúa Korsgaard, se aleja de la dicotomía empirista entre razón e inclinación (modelo de combate), puesto que actuar de acuerdo con razones no se opone a actuar por inclinación. Las inclinaciones y las razones cumplen papeles distintos en la explicación de la acción. Las primeras nos indican que hay un incentivo presente, mientras que las segundas sirven para decidir si actuar o no de acuerdo con la inclinación. La razón no funciona como el instinto animal. La idea es que la inclinación presenta una propuesta [*proposal*] a la razón, y esta sería la encargada de decidir si asiente o no a la propuesta. En el modelo constitucional kantiano, entonces, la razón vendría a ser la facultad que asiente o rechaza los incentivos que se le proponen.

La razón ejerce su legislación a través de leyes, pero no se trata de leyes cualesquiera. ¿Cómo decide la razón si asiente o no a una propuesta? La respuesta a esta pregunta explicita el compromiso de Korsgaard con la filosofía moral kantiana. Según esta autora, la razón asiente o rechaza una propuesta a través del imperativo categórico. Este imperativo no nos impone ningún contenido. El único requerimiento que él plantea es que "nuestra elección ha tener la forma de una ley universal" (Korsgaard, 1996, p. 98). El imperativo categórico sería la forma en la que el agente toma distancia reflexiva con respecto a las propuestas que provienen de la inclinación. Korsgaard afirma que cuando las máximas de la inclinación se nos presentan como propuestas, siempre podemos considerar, como agentes libres y autónomos, si

queremos que se conviertan en una ley para nuestra voluntad. En tanto que somos agentes libres, estas leyes no pueden ser externas a la voluntad pues en ese caso esta sería heterónoma. Una voluntad autónoma debe darse a sí mismas las leyes. Y, como sabemos, la ley que cumple esta condición de autonomía es el imperativo categórico (§3.4).

A partir de este proceso reflexivo que implica el imperativo categórico, Korsgaard refuerza el argumento según el cual el modelo de la agencia kantiano se ajusta al modelo constitucional. "La inclinación presenta la propuesta; la razón decide si actúa de acuerdo con ella o no, y la decisión toma la forma de un acto legislativo. Este es claramente el Modelo Constitucional" (Korsgaard, 1999, p. 12). El agente kantiano, entendido bajo el modelo constitucional, es capaz de ejercer una legislación racional sobre sus inclinaciones. Esto hace que la agencia no se reduzca a un conflicto entre la inclinación y la razón. A diferencia de lo que sucede en el modelo de combate, lo que propone Korsgaard es que la agencia es un ejercicio de constitución racional. Este ejercicio se manifiesta cuando el agente ejecuta su capacidad racional guiado por el imperativo categórico a través del cual puede asentir y elegir las distintas propuestas o las distintas máximas de acción.

Con lo anterior quiero hacer notar que este ejercicio de constitución de la agencia es, en sentido estricto, un ejercicio de *auto*-constitución: es el agente quien, a través de sus decisiones guiadas por el imperativo categórico, se constituye en autor de sus acciones (la auto-constitución, de hecho, está estrechamente ligada a la autonomía de la voluntad). Es en la medida en que las acciones de un agente son una expresión de este ejercicio racional, que este se convierte en el autor de su constitución, o en otras palabras, el agente se auto-constituye.[70]

---

[70] El modelo constitucional conduce a lo que Korsgaard denomina la *paradoja de la auto-constitución*: "¿Cómo puedes auto-constituirte, crearte a ti mismo, a menos que ya estés ahí?" (Korsgaard, 2009, p. 20). Si ser-agente no es una condición dada sino que es una tarea por realizar, hace falta entender *quién* lleva a cabo esta tarea. La solución a esta paradoja radica en enteder el concepto de *auto-constitución:* quien realiza la constitución es el agente mismo, y la auto-constitución se realiza en cada una de las acciones que él lleva a cabo. Cada acción que realiza el agente, como he explicado, es un ejercicio racional de acuerdo con principios que lo unifican o lo constituyen como un todo unificado, luego en cada acción se instancia la constitución de la agencia. Y, como veremos, estos principios surgen o tienen su fuente en las identidades prácticas (§5.2.3). De manera que existe algo previo a la agencia, la identidad práctica, que le sirve al ser racional para guiar su elección. De modo que, cuando

En conclusión, la consecuencia de leer la agencia kantiana en términos del modelo constitucional es que la constitución del agente se logra cuando este ejecuta su capacidad racional sobre la base del imperativo categórico. Esta auto-constitución es lo que unifica o integra al agente como un todo. "El trabajo de alcanzar la unidad psíquica […] es lo que llamo auto-constitución" (Korsgaard, 2009, p. 7). Para fortalecer esta conclusión, sin embargo, habría que especificar con más precisión por qué el imperativo categórico unifica al agente como un todo integrado.

5.2.3 Identidades prácticas

Para explicar con más precisión en qué consiste la unificación del agente es necesario hacer un rodeo por la noción de identidad práctica. Esta noción se desprende del hecho de que los seres humanos somos auto-conscientes. La capacidad de tomar una distancia reflexiva con respecto a nuestros deseos, intenciones, inclinaciones, etc., necesariamente nos impela a asentirlas o a rechazarlas. Y este proceso de asentimiento o de rechazo consiste en preguntarnos por qué asentimos o rechazamos tal deseo, inclinación o intención. Este modo de ejercer la capacidad reflexiva nos obliga a tener una concepción de nosotros mismos (Korsgaard, 1996, p. 100), es decir, en la medida en que somos auto-conscientes podemos tomar distancia con respecto de nosotros mismos, y de esa manera podemos hacernos una imagen o concebirnos como siendo de alguna manera específica.

La imagen o concepción que nos hacemos de nosotros mismos sería la representación de aquellos principios y normas que adoptamos cuando, por ejemplo, nos embarcamos en el ejercicio de deliberar sobre un curso de acción o sobre una propuesta que provenga de nuestros incentivos o deseos. Por ejemplo, si me concibo a mi mismo como una persona abstemia, eso significa que rechazaré cursos de acción en los que podría beber alcohol. O si

---

actúa, la persona lo hace dentro del marco normativo de una identidad práctica; no delibera en el vacío. Y al deliberar de acuerdo con estos principios es cuando el ser reacional está constituyéndose como agente. "Lo que es ser una persona, o un agente racional, es estar comprometido en la actividad de constantemente llegar a ser una persona" (Korsgaard, 2009, p. 42).

me concibo como un padre responsable, rechazaré todo aquello que pueda afectar a mis hijos. Lo que busco decir es que como seres auto-conscientes no deliberamos únicamente a través del imperativo categórico. Además, somos capaces de deliberar, asentir y elegir a partir de otros principios, normas, leyes, deberes, que, al igual que el imperativo moral, tienen un peso normativo para nosotros. El conjunto estos elementos normativos es lo que conforman una identidad práctica.

Las identidades prácticas tienen fuerza normativa. Cuando un agente se identifica con una identidad práctica surgen compromisos, obligaciones y exigencias. Por ejemplo, se exige valor a los soldados, confidencia a los psiquiatras y médicos, amor y cuidado a los padres. Y esto vale también para las distintas identidades prácticas sociales: ser un cristiano practicante exige ciertas acciones y compromisos que son distintos en el caso del creyente musulmán. Ser un vegano supone unas exigencias que no tiene una persona no vegana, etc. La noción de identidad práctica, de esta manera, está necesariamente ligada a exigencias de normatividad. Las identidades prácticas que tengamos son fuente de deberes y obligaciones específicas. "Todas estas identidades dan lugar a razones y obligaciones. Tus razones expresan tu identidad, tu naturaleza; tus obligaciones nacen de lo que identidad prohíbe" (Korsgaard, 1996, p. 101). Y es a partir de estas obligaciones que se configura la deliberación práctica de un agente. De hecho, la deliberación práctica no se hace en el vacío, deliberamos sobre la base de un marco normativo que está determinado por alguna o varias de nuestras identidades prácticas. La deliberación que lleva a cabo un vegano no es la misma que llevaría a cabo un no-vegano, o un médico no deliberará igual que un soldado. Ni una madre igual que un juez. Cada uno de ellos respondería a principios normativos diferentes porque cada uno de ellos se identifica con identidades prácticas diferentes.

> La concepción de la propia identidad que se investiga aquí no es teórica, una aproximación sobre lo que eres como si se tratara de un hecho científico ineludible. Es mejor entenderla como una descripción bajo la cual te valoras a ti mismo, una descripción bajo la cual encuentras tu vida digna de ser vivida y tus acciones dignas de ser llevadas a cabo. Llamaré a esto una concepción de tu identidad práctica. La identidad práctica es un asunto complejo y para las personas en general habrá una

> confusión sobre estas concepciones. Tú eres un ser humano, una mujer o un hombre, un feligrés de cierta religión, un miembro de cierto grupo étnico, un miembro de cierta profesión, el amante o el amigo de alguien, y así sucesivamente. (Korsgaard, 1996, p. 101)

Con respecto a la noción de identidad práctica, hay tres ideas que me interesa resaltar y que desarrollaré a continuación: (i) el carácter contingente de la mayoría de las identidades prácticas, (ii) la noción de identidad práctica humana y (iii) la identidad práctica moral.

a. Contingencia

Una característica de las identidades prácticas es que la mayoría de ellas son contingentes. Hay un primer sentido en que nuestras identidades son contingentes: muchas de nuestras ellas se adquieren por simple fortuna y al margen de nuestro control: el lugar donde nacimos, la condición socio-económica, el país al que pertenecemos, la religión de nuestros padres o cuidadores, etc. (2009, p. 23). Otras identidades dependen en cierta medida de nosotros, como la profesión que elegimos o la persona con quien nos casamos. Así que muchas de ellas, aunque no todas, son adquiridas al margen de nuestra voluntad[71].

Hay un segundo sentido en el que las identidades prácticas son contingentes: todas son fuente de obligaciones y deberes, solo si el agente reconoce y se identifica con una determinada identidad. La identificación con una identidad práctica es lo que la hace operativa en un sentido práctico. Por ejemplo, ser hijo es una identidad práctica que tengo y a la cual no puedo renunciar. Ella está socialmente configurada, porque la relación padres-hijo depende de múltiples factores socio-culturales; y al mismo tiempo es también una identidad irrenunciable, pues tiene bases naturales. Aun así, es posible determinar si las obligaciones y deberes que se desprenden de esta identidad constituyen o no una razón para actuar. Aunque no pueda elegir tener una identidad práctica, sí se puede elegir adoptarla como fuente de razones. De esta manera, las identidades son contingentes porque no estamos obligados

[71] Sobre el concepto de fortuna moral, ver Williams (1981c); y específicamente sobre la noción de fortuna moral constitutiva, ver Nagel (1979a).

necesariamente a comprometernos con ellas. Lo que Korsgaard señala es que un agente puede distanciarse o incluso renunciar a la mayoría de sus identidades prácticas. En otras palabras, es posible dejar de identificarse con la mayoría de ellas. Incluso, dice Korsgaard, es posible que surjan conflictos entre las distintas identidades que nos obliguen a renunciar a una de ellas. Si una persona ya no se identifica con cierta identidad práctica, básicamente esta identidad ha perdido su autoridad normativa sobre las acciones y sobre las deliberaciones que ella lleva a cabo.

> [S]i tratas estas identidades como fuente de tus razones y obligaciones es algo que depende de ti. Si continúas adoptando las razones que la identidad te presenta, y cumples las obligaciones que te impone, entonces eres tú [quien las adopta]. (Korsgaard, 2009, p. 23ss)

En el sentido opuesto, las identidades con las que nos identificamos pueden llegar a ser normativamente muy poderosas, pues ellas son "descripciones bajo las cuales valoras tu vida y encuentras tu vida digna de ser vivida y tus acciones dignas de ser llevadas a cabo" (Korsgaard, 2009, p. 20). Aunque la mayoría de nuestras identidades sean contingentes, cuando nos identificamos fielmente con ellas nos resultan una "fuente de leyes inviolables absolutas" (*ibid.*). El agente puede tener algunas identidades prácticas que son muy importantes para él y para la concepción que tiene de sí mismo. Su identidad personal puede estar constituida, en un sentido muy profundo, por identidades prácticas muy específicas. Por esta razón, las obligaciones y compromisos creados a partir de ellas resultan ser apremiantes, obligatorias y prácticamente incondicionales.

Cuando una persona tiene un compromiso incondicional con alguna de sus identidades prácticas, las obligaciones que de allí se desprenden también resultan incondicionales y fundamentales para la concepción que tiene la persona sobre sí misma. Esta idea es central para esta investigación y será desarrollada en §6.2.2. Por ahora, lo que me interesa resaltar es que las identidades prácticas incondicionales son parte fundamental de la vida de una persona, pero eso no significa que ella no pueda renunciar a ellas, después de todo, esas identidades son contingentes. No obstante, como la misma Korsgaard reconoce, puede suceder que renunciar a una identidad práctica incondicional suponga la muerte práctica

(§4.2.1e):

> Son las concepciones de nosotros mismos que más nos importan las que dan lugar a obligaciones incondicionales. Violarlas significa perder tu integridad y tu identidad, y dejar de ser quien eres. Es decir, dejas de ser capaz de pensarte a ti mismo bajo la descripción bajo la cual te valoras a ti mismo y encuentras tu vida valiosa y tus acciones dignas de ser realizadas. Para todos los propósitos prácticos es una muerte o peor que la muerte. (Korsgaard, 1996, p. 102)

b. Identidad práctica humana

¿Por qué, si son contingentes, puede haber identidades prácticas tan importantes para una persona? La respuesta general a esta pregunta es que existe una jerarquía de identidades. Y esta jerarquía depende del grado con que una persona se identifique con tal o cual identidad práctica. Las identidades con las que alguien más se identifique tendrán más peso normativo en comparación con aquellas identidades con las que no haya una identidad tan fuerte.

Ahora bien, es verdad que la mayoría de identidades prácticas son contingentes, "[p]ero lo que no es contingente es que debamos estar gobernados por *alguna* concepción de tu identidad práctica" (Korsgaard, 1996, p. 120; énfasis en el original). En otras palabras, una persona puede renunciar a tal o cual identidad, pero no puede renunciar a todas. No puede hacerlo porque sin ninguna identidad práctica ella no tendría ningún principio, norma o ley que guíe su acción; es decir, sin identidad práctica alguna, la persona no tendría ningún marco normativo para actuar y, por tanto, no tendría ninguna razón para hacer o dejar de hacer algo. "Es necesario tener alguna concepción de tu identidad práctica, ya que sin ella no habría ninguna razón para actuar" (*ibid.*). Con esto busco decir que sin alguna identidad práctica a la persona le daría lo mismo hacer o no hacer determinadas cosas; su deliberación carecería de cualquier principio práctico. No tener ninguna identidad significaría perder el control de uno mismo, significaría que nos daría lo mismo tener o no una razón para llevar a cabo una acción (*ibid.* p. 121).

Carecer de toda identidad práctica no solo paraliza o hace imposible la acción, el problema

es más profundo. Resulta que 'actuar' y 'actuar con base en razones' es lo mismo, porque la acción, según el modelo constitucional, es el resultado de un ejercicio racional en el que el agente elije según principios; y estos principios son las razones que él tiene para actuar que surgen de las exigencias normativas de sus distintas identidades prácticas[72]. Con esto se ve entonces que si una persona renuncia a toda identidad práctica, no tendría ninguna razón para actuar. Y es que si la persona carece de razones, sus movimientos no podrán contar siquiera como acciones.

La acción, repito, requiere siempre que haya una razón para llevarla a cabo. Esta concepción de la acción, dentro del modelo constitucional, supone que cuando el agente actúa tome *distancia reflexiva* con respecto a sus deseos, incentivos, inclinaciones, etc., y decida de acuerdo a principios. Esto es lo que Korsgaard tiene en mente cuando dice que el agente "toma el control de sus movimientos" y decida según los principios de su identidad práctica (Korsgaard, 2009, p. xi). De modo que si renunciamos a la posibilidad de actuar con base en razones, es decir, si renunciamos a cualquier principio o norma que nos impongan nuestras distintas identidades prácticas, estamos renunciando también a la característica humana de autoconsciencia, pues estaríamos anulando la capacidad que tenemos de tomar distancia reflexiva.

Lo que me interesa resaltar es que existe una estrecha relación entre la noción de identidad práctica y la condición humana de auto-consciencia. La identidad práctica solo es concebible en la medida en que somos seres autoconscientes, o en tono trascendental, la auto-consciencia es condición de posibilidad de la identidad práctica, porque sin la distancia reflexiva no podríamos siquiera concebirnos como seres con identidad práctica alguna. De manera que la idea misma de identidad práctica solo es posible por la auto-consciencia. Y ya que la auto-consciencia es intrínseca a nuestra condición de humanidad, resulta que no podemos tener

[72] "Cuando un impulso –es decir, un deseo– se presenta ante nosotros, nos preguntamos si este puede ser una *razón*. Nosotros respondemos esta pregunta examinando si la máxima de acción de actuar según el deseo puede ser querida como un ley por un ser con la identidad en cuestión. Si puede ser deseado como una ley, esta máxima es una *razón*, ya que tiene una estructura intrínsecamente normativa. Si no puede ser deseada como ley, debemos rechazarla, y en ese caso tenemos la obligación [de no actuar de acuerdo al deseo]" (Korsgaard, 1996, p. 113; énfasis añadido).

ninguna concepción de identidad práctica sin concebirnos al mismo tiempo como seres humanos. En otras palabras, la idea misma de identidad práctica tiene como condición de posibilidad la identidad humana.

La identidad humana se convierte entonces en un aspecto constitutivo de la acción y de la agencia porque en virtud de ella tiene sentido la idea misma de 'razón para actuar'. Esto explica por qué las personas estamos obligadas a tratar nuestra humanidad como una forma de identidad práctica normativa, es decir, la posibilidad de tener una identidad práctica depende de que la persona valore normativamente la condición de humanidad que le es propia (Korsgaard, 1996, p. 121). Con este razonamiento, Korsgaard estaría mostrando que una persona no puede no tener una identidad práctica ni, por tanto, no puede no tener razones para actuar, a menos que su identidad humana tenga valor o peso normativo.

En breve, la idea de Korsgaard es de corte trascendental: la condición de posibilidad de ser-agente es actuar con base en razones. Y tener razones solo es posible gracias a que tenemos alguna identidad práctica; sin ninguna identidad práctica no tendríamos razón alguna para actuar. Y, sigue el argumento: la condición de posibilidad de la identidad práctica es, a su vez, la condición humana de autoconsciencia, porque es en virtud de la autoconsciencia que se hace inteligible la idea misma de tomar distancia reflexiva para poder actuar con base en razones. Dado que la autoconsciencia es un rasgo propio de la condición humana, entonces una persona no puede tener identidad práctica alguna a menos que reconozca también su identidad humana (su capacidad autoconsciente para actuar con base en razones).

Llegar a este punto tiene una importante consecuencia normativa: si la identidad humana es condición de posiblidad de cualquier otra identidad práctica, entonces la identidad humana adquiere un valor normativo para determinar qué cuenta como una acción. Con esto busco decir que una acción solo puede ser aquella que deriva de la capacidad humana de dar y recibir razones, propia de la identidad humana. Si una persona no actúa de acuerdo con razones, ella no está realizando una acción, no está operando como agente; tal vez esté siendo movido por fuerzas internas o externas, o lo que hace tal vez sea un evento causado por una parte del cuerpo. El punto, insisto, es que la norma de la acción y de la agencia sería la

capacidad de actuar con base en razones, y que esto es propio de la identida humana. Este es el sentido en el que esta identidad tiene carácter normativo. La conclusión que deriva Korsgaard de acá es que estamos normativamente ligados a la identidad humana porque solo así tiene sentido la idea de ser agentes que actuamos de acuerdo con razones. O dicho de otra manera, la fuente de la normatividad de la identidad humana radica en que esta es constitutiva de nuestra condición de agentes. Sin identidad humana no podríamos darle sentido a la agencia.

c. Identidad moral

La implicación de comprometerse con la normatividad de la identidad humana es que nos ubica en el territorio de la moral. Con esto busco decir que el reconocimiento de mi identidad humana es indesligable del reconocimiento de la identidad humana de los otros seres humanos. Al reconocer la condición humana de los otros, se entiende en qué sentido la identidad humana es una identidad moral: dado que mi propia identidad humana tiene valor para mí, y dado que puedo reconocer esta identidad humana en otros seres humanos, necesariamente debo reconocer que la condición humana (la propia y la ajena) tienen un valor que no puedo simplemente omitir. La idea de Korsgaard es que si valoramos la propia humanidad, necesariamente debemos valorar la humanidad de los otros seres humanos (Korsgaard, 1996, p. 121); o en otras palabras, el valor que otorgo a mi humanidad me lleva a reconocer el valor de la humanidad de los otros seres humanos. La identidad humana, entonces, nos ubica en un plano moral: no puedo concebir ni valorar mi propia humanidad sin concebir esa misma condición en los otros.

Con esta idea de claro cuño kantiano, Korsgaard está trazando un puente entre la agencia y la moralidad. La identidad humana es como una bisagra que articula la teoría de la agencia con la teoría moral. Esta afirmación no es obvia y será ampliada en la sección siguiente. Lo cierto es que si Korsgaard tiene razón, si es verdad que al valorarnos a nosotros mismos como seres humanos valoramos al mismo tiempo la humanidad de los otros, entonces la identidad humana es ya una identidad moral (Korsgaard, 1996, p. 121).

Con esto claro, es posible dar cuenta la prioridad normativa de la identidad humana-moral. Korsgaard sostiene que la identidad humana-moral soporta [*stands behind*] a las demás identidades prácticas contingentes, pues, como vimos, la importancia de estas últimas deriva del valor de nuestra identidad en tanto que seres humanos. "De esta forma todos los valores dependen del valor de la humanidad; otras formas de identidad práctica importan porque la humanidad requiere de ellas" (Korsgaard, 1996, p. 121). Existe, pues, una suerte de primacía o prioridad normativa[73] de la identidad humana-moral que se explica porque cualquier identidad práctica contingente depende de esta.

> La mayoría del tiempo, nuestras razones para la acción surgen de nuestras identidades más contingentes y locales. Pero parte de la fuerza normativa de estas razones nace del valor que nos otorgamos a nosotros mismos como seres humanos que necesitamos tales identidades. De esta manera todo el valor depende del valor de la humanidad; otras formas de identidad práctica importan en parte porque la humanidad las requiere. La identidad moral y las obligaciones que conlleva son por consiguiente ineludibles y dominantes. No toda identidad práctica es contingente o relativa, después de todo: la identidad moral es necesaria. (Korsgaard, 1996, p. 121s)

5.2.4 Razones públicas y razones privadas

He analizado el concepto de identidad práctica y el sentido en que esta identidad es una identidad moral. Sin embargo, esto no es suficiente para explicar la forma en la que, según

[73] El concepto de *prioridad normativa* está en estrecha relación con el concepto de *prioridad deliberativa:* "Una consideración tiene una alta prioridad deliberativa para nosotros si le damos una gran importancia frente a otras consideraciones en nuestras deliberaciones […]. Si algo te resulta importante, entonces eso afectará tu vida de una manera u otra, y asimismo afectará tus deliberaciones" (Williams, 2006, p. 183). Una de las consecuencias de que algo tenga prioridad deliberativa es que tendrá igualmente prioridad normativa. Esto, evidentemente, se refiere a la jerarquía normativa que tiene una deliberación con respecto a otro tipo de consideración práctica. Williams, por su parte, considera que la tradición filosófica le ha otorgado a la moralidad una prioridad normativa *a priori* que es muy cuestionable. La prioridad normativa o deliberativa, dice Williams, está asociada a aquello que más nos importa y no tanto a aquello que debemos hacer o aquello que estamos obligados a hacer en terminos morales. En el capítulo 6 ampliaré esta idea.

Korsgaard, el agente kantiano está unificado o integrado. Para ello aún resta entender cómo se integran las identidades particulares no-morales con la identidad moral. La forma en la que esta autora explica la relación entre estos dos tipos de identidades es a través de la distinción entre razones privadas y razones públicas.

> Nuestras razones deben ser lo que denomino razones públicas, razones cuya fuerza normativa se extiende más allá de los límites entre las personas. Las razones públicas son aproximadamente lo mismo que las razones objetivas o neutrales con respecto al agente. Estas pueden ser contrastadas con lo que denomino razones privadas —razones subjetivas o relativas al agente. Una razón privada es una razón cuya fuerza normativa es privada, en el sentido de que pertenece solo a una persona. (Korsgaard, 2009, p. 191)

El siguiente ejemplo de Korsgaard puede ilustrar esta distinción. Supongamos que un estudiante y su profesora están acordando una reunión. La profesora propone que se vean después de clase, pero el alumno replica que a esa hora no puede porque tiene otra clase. En este caso, tendrán que buscar otro horario. La reunión es una actividad compartida, así que un horario que no favorezca al estudiante o a la profesora no será favorable para ambos. "Para llevar a cabo una acción compartida, cada uno de nosotros ha de adoptar las razones del otro como propias" (Korsgaard, 2009, p. 192). En este sentido, la excusa del estudiante tiene fuerza normativa porque se trata de una razón pública. En cambio, si el estudiante replicara que a esa hora no le va bien la reunión porque, digamos, estará durmiendo hasta tarde, tendríamos un escenario en que su razón privada no podría ser tenida en cuenta como una razón para cambiar el horario.

Las razones públicas tienen una fuerza normativa que ha de ser válida objetivamente, es decir, han de ser válidas para otras personas. "En la concepción pública de las razones un requerimiento de universalidad [*universalizability*] me compromete con que si yo tengo una razón para hacer la acción-A en la circunstancia-C, debo ser capaz de *querer* que tú debas hacer la acción-A en la circunstancia-C […]" (Korsgaard, 2009, p. 192; énfasis en el original). Las razones privadas, por su parte, también pueden tener fuerza normativa, pero solo en un sentido individual o privado. Una razón privada puede explicar por qué hago tal o cual

actividad (p.e. escalar una montaña, coleccionar estampillas, escribir libros o realizar donaciones). Además, las razones privadas puede que tengan fuerza motivacional suficiente para llevar a cabo ciertas acciones. Sin embargo, este tipo de razones no necesariamente cuentan como razones para otras personas.

Existe entonces una asimetría normativa entre las razones privadas y las públicas. No obstante, Korsgaard indica acertadamente que los seres humanos somos "animales sociales" que estamos ya situados en un espacio público, de manera que las razones privadas propias y las razones privadas de las otras personas son percibidas como públicas (*cf.* Corbí, 2012, p. 111). De ahí que no exista una brecha insalvable entre razones públicas y privadas. No hay ninguna inconsistencia entre tener razones privadas que sean también públicas. En otras palabra: las razones privadas, dadas ciertas condiciones, pueden perfectamente tener normatividad pública.

> [L]a primera persona no necesita ser la primera persona singular: el hecho de que una razón tenga que ser mía en esta forma no impide que sea *nuestra*. Así entendida, la categoría de mis razones no excluye una identificación entre mis razones y las razones de otros. (Korsgaard, 2009, p. 198; énfasis en el original)

Nuestra naturaleza social no elimina el hecho de que yo pueda tener razones privadas, pero sí enriquece el espacio de las razones, pues abre la posibilidad de tener en cuenta la interacción con otras personas. Si atiendo a esta forma de interacción social, entonces tanto mis razones para actuar, como las razones de los otros se convierten en posibles razones en un espacio público. De esta manera, las razones privadas que puedan surgir de mi identidad práctica están en un espacio en el que también habitan las razones de los otros.

> Yo [...] debo actuar con mi cuerpo. Pero podría también interactuar con otras personas, y cuando lo hago, entonces sus razones, tanto como las mías, llegan a ser como si fueran incentivos en el proceso deliberativo que emprendemos juntos, fuentes para la construcción de nuestras razones compartidas. […] La categoría de *mis* razones no excluye una identificación entre mis razones y las razones de otros. No necesito legislar solo, y siendo así, no necesito considerar los incentivos que

> surgen en el curso natural de mi vida individual como las únicas fuentes posibles de razones para mí. (Korsgaard, 2009, p. 199ss; énfasis en el original)

Es posible que la relación entre las razones públicas y las razones privadas no resulte tan sencilla como piensa Korsgaard. Un defensor de las razones privadas —como Hume (1976) o como Williams (1981a)— necesitaría un argumento adicional para entender por qué un agente debería tener en cuenta las razones de otras personas como razones propias. Korsgaard y un internalista como Williams estarían de acuerdo en que los seres humanos compartimos un espacio público, pero esto no es suficiente para exigir que las razones públicas (externas) o las razones privadas de otros tengan fuerza normativa.

Korsgaard acepta que la normatividad de las razones públicas no puede derivarse de escenarios donde haya un fin compartido (acción colectiva) o donde hay un acuerdo o contrato (contractualismo). En el primer caso, porque si dos agentes tienen un fin compartido, las razones de la otra persona solo son instrumentos que se usan para alcanzar el fin compartido; de modo que las razones del otro no son en sí mismas razones, son sólo medios para satisfacer un interés personal. En el segundo caso, cuando alcanzamos un acuerdo, es posible que se tengan en cuenta las razones de otros solo de manera instrumental para favorecer un acuerdo que beneficia a ambas partes; sin embargo, en este escenario no hay garantía de normatividad porque siempre estará presente el fantasma del oportunista (*free rider)* ¿Por qué querría alguien cumplir un acuerdo, si puede obtener más beneficio al incumplirlo?.

En estos dos escenarios —el del fin compartido y el del acuerdo— se presupone que las razones, primero son privadas y luego, dado cierto interés o tras un acuerdo, llegan a ser compartidas. Korsgaard, en contraste, adopta una salida kantiana: "nosotros tenemos que tratar nuestras razones como compartibles [*shareable*] en primer lugar, con el objetivo de hacer acuerdos. No podemos hacer un acuerdo para tratar nuestras razones como compartibles" (Korsgaard, 2009, p. 201; énfasis añadido). En los escenarios en los que se tienen en cuenta las razones del otro porque me sirven para un propósito, en realidad no estoy tratando las razones del otro como razones, "porque si mi postura hacia ellas es instrumental yo estoy considerándolas necesariamente como instrumentos" (*ibid.*). Para Korsgaard, tratar una

razón como teniendo una fuerza normativa significa que debo aceptar que las razones tienen fuerza normativa "simplemente porque *son* tus razones" (*ibid.*).

a. La naturaleza pública de las razones

El argumento de Korsgaard sirve para explicitar bajo qué condiciones las razones de otra persona tienen fuerza normativa. Sin embargo, con esto no queda claro por qué esas razones deberían ser aceptadas. Aun dado esto, para Korsgaard, existe todavía una forma de mostrar que las razones privadas deben poder estar sometidas a la normatividad pública: un agente racional que solo tenga razones privadas está condicionado a que sus razones cuenten como razones para él mismo, es decir, está condicionado a que sus razones privadas sean normativas en un espacio que incluya, al menos, dos puntos de vista: el suyo propio y el de un yo-hipotético. En otras palabras, una razón privada solo cuenta como una razón, si puede ser asentida por al menos el yo-hipotético. De manera que cualquier razón privada ya exige un espacio de razones no solipsista. Para que algo cuente como mi razón, tengo que estar de acuerdo conmigo mismo. Y esto equivale a ubicar mi razón en un espacio público de razones.

Para Korsgaard "[H]ay una persona con quien incluso el razonador privado más estricto debe interactuar en la forma en que la teoría kantiana requiere. Y es él mismo" (Korsgaard, 2009, p. 202). Para esta autora, la relación que tiene una persona con otros es semejante a la relación que ella tiene consigo misma. No hay que interpretar esta idea como si Korsgaard quisiera hacer prescindible la interacción social. Su punto, más bien, es que la relación que una persona tiene consigo misma es más íntima y más constante que la relación que existe entre personas; y además, la relación con uno mismo es inevitable, despues de todo, nadie puede evitar el trato conmigo mismo. El solipsismo normativo del razonador privado es, en esta medida, una ilusión. A fin de cuentas una razón privada debe poder ser validada no porque es mía, sino porque cuenta con el asentimiento de, al menos, otro –el yo-hipotético– que funciona como un agente racional que evalúa la validez de mis razones privadas.

> Constituir tu propia agencia es un asunto de elegir solo aquellas razones que tú puedas compartir contigo mismo. Por esto tú tienes que desear universalmente,

> porque la razón con la que actúas ahora, la ley que creas para ti mismo ahora, debe ser tal que puedas desearla nuevamente más adelante, sin importar lo que suceda, a menos que tú veas que hay una buena razón para cambiarla. (Korsgaard, 2009, p. 202ss; énfasis en el original)

Esto significa que si mi razón es solo válida para mí en una situación muy particular y muy específica, no habría garantía de que mi yo futuro estuviera de acuerdo con ella. De ahí que mi razón debe tener una validez normativa que valga en el espacio público de razones, es decir, que valga para mi yo-hipotético y también para otros seres racionales. Dicho de otra manera, una razón privada que yo pueda compartir conmigo mismo es una razón que está en el espacio público y, por lo tanto, es en realidad una razón pública.

Llegados a este punto resulta que una razón que yo pueda compartir conmigo mismo y con los demás debe ser expresada en términos de un principio universal, un principio que sea válido para todos. Al cumplir con estas características, una razón pública adquiere la categoría de ley universal, pues cumple con la condición de ser válida para todos. Esta afirmación es el corazón kantiano de la propuesta de Korsgaard. Los principios válidos para todos los seres racionales no pueden ser entonces razones privadas. Cuando una persona legisla en el espacio de las razones públicas en realidad está en el terreno de las leyes universales. Lo anterior significa que una razón privada que no pueda ser ubicada en el espacio público no es una razón con la que yo mismo pueda estar de acuerdo.

> Las razones que legislas cuando deseas la ley tienen que ser públicas, es decir, tienen que tener una fuerza normativa que pueda ser compartida por todos los seres racionales, porque actuar es interactuar contigo mismo —contigo en otros momentos y en otras situaciones posibles. Tienes que hacer leyes para ti mismo. Y a menos que las leyes que hagas ahora te obliguen en otros momentos y en otras situaciones, y a menos que las leyes que tú sabes que harás en otros momentos y en otras situaciones te obliguen ahora, ellas no te mantendrán reunido en una unidad después de todo. Para lograr esto, las razones que tú legisles deben ser públicas. Las leyes tienen que ser leyes para todos los seres racionales, leyes cuya fuerza normativa pueda ser compartida. (Korsgaard, 2009, p. 214)

Esta conclusión es bastante favorable para el modelo constitucional que Korsgaard propone: las razones privadas, aquellas que nacen de mis identidades prácticas personales, han de cumplir el requisito de que sean asentidas universalmente. Si no fuera así, el agente mismo no podría estar de acuerdo con sus razones privadas. "Así, todo agente racional debe desear de acuerdo con una ley universal […]" (Korsgaard, 2009, p. 214).

De hecho, según Korsgaard, no estar dispuesto a adoptar la ley universal equivale a una falta de auto-respeto[74]. Guiarse por razones privadas que no puedan ser universalizables es desatenderse o menospreciarse a sí mismo, pues una razón privada particular no tiene en cuenta al yo posible en el futuro. Esos yoes-posibles esperan que el principio subjetivo que rija la acción en este momento sea válido también en cualquier otra circunstancia. Y aquí ya puede entreverse, de modo general, que esta fractura entre el yo presente y el yo futuro es una forma de desunificar o desintegrar al agente. Por el contrario, el asentimiento compartido por el yo presente y el yo futuro sería lo que unifica o integra al yo a través de una razón *compatida*. Aunque volveré sobre esta idea más adelante, es posible intuir por ahora que la integridad o unidad del agente parece radicar en la existencia de razones públicas que puedan ser compartidos por el yo a lo largo del tiempo.

b. Ley moral

Hasta este punto del argumento, Korsgaard ha mostrado que una razón solo puede ser legítima si es válida en un terreno público, es decir, si se constituye como una razón pública y universal para todos los seres racionales (incluyendo a mi yo futuro). ¿Cómo establecer razones públicas que sean válidas universalmente? El procedimiento para establecer si una razón privada puede ser asentida universalmente no es otro que el imperativo categórico. Es a partir del imperativo que es posible verificar [*test*] si una razón privada está justificada públicamente. Esto es lo que garantiza la imparcialidad del procedimiento y también que los resultados de la deliberación puedan ser reconocidos públicamente, es decir, que puedan ser

[74] Guiarse por razones privadas particulares es, según Korsgaard, no gobernarse a sí mismo por una ley que nazca de la propia voluntad. "Y si no deseas tu máxima como ley universal, entonces careces de auto-respeto. Más precisamente, ese hecho es lo mismo que el hecho de carecer de auto-respeto, ya que no estás gobernándote a ti mismo por la ley de tu propia voluntad" (Korsgaard, 2009, p. 203).

razones públicas. De manera que una razón privada, cuando es sometida al escrutinio normativo, en realidad está siendo sometida al criterio del imperativo categórico. Decir que asiento universalmente ante una razón privada es lo mismo que decir que mi razón privada se adecúa a esa forma de ser del imperativo categórico.

Sin embargo, el procedimiento del imperativo categórico es formal y no dice nada respecto al contenido de las razones. Korsgaard está sugiriendo que el imperativo categórico es un procedimiento meramente formal que carece de contenido (*cf.* Korsgaard, 1996, p. 98ss). ¿Existe algún tipo de contenido normativo para lo que cuenta como una razón pública?

Korsgaard es consciente de los límites formales del imperativo categórico. No obstante, ella está interesada no solo en mostrar la normatividad de las razones públicas, sino en mostrar que la normatividad obedece a un criterio moral; es decir, pretende dotar de contenido a las razones públicas. Para este objetivo, y a diferencia de Kant, la autora distingue entre el imperativo categórico y la ley moral:

> Haré una distinción que Kant no hace. Voy a llamar a la ley de actuar por máximas de las que puedas querer que sean leyes 'el imperativo categórico'. Y voy a diferenciarla de lo que llamaré 'la ley moral'. La ley moral, en el sistema kantiano, es la ley de aquello que Kant llama el reino de los fines, la república de todos los seres racionales. La ley moral nos dice que actuemos solamente según máximas sobre las cuales todos los seres racionales, en un sistema factible cooperativamente, puedan estar conjuntamente de acuerdo. (Korsgaard, 1996, p. 99)

El imperativo categórico solo necesita atenerse al criterio de universalización de la máxima; solo tiene el propósito de que el agente actúe de acuerdo con máximas que él pueda querer que sean leyes universales. La ley moral, por su parte, tiene una pretensión normativa más fuerte, pues ya no se trata de una ley que la voluntad del agente pueda querer universalmente, sino de una ley a la que todos los miembros del reino de los fines puedan asentir. La ley moral es la ley del reino de los fines y como tal supone la comunidad de todos los seres racionales.

Al introducir la noción de reino de fines, Korsgaard está tratando de dar contenido al formalismo kantiano. El reino de los fines es el nombre para designar la "conjunción

sistemática de seres racionales" (*F* 4:433) que tienen valor por sí mismos y por tanto cuentan como fines en sí mismos. Ahora bien, con la idea del reino de los fines se quiere dar a entender, no solo que las razones públicas requieren del asentimiento universal, sino además que este asentimiento es propio de los seres humanos racionales que son miembros de este reino. En otras palabras, para que las razones sean válidas públicamente para todos los seres racionales habrá que presuponer una comunidad de seres que son capaces de dar y recibir razones en un espacio público.

Ahora bien, en este punto es importante recordar que esta capacidad de dar y recibir razones se deriva de la condición de autoconsciencia propia de los seres humanos. En otras palabras, la capacidad de dar y recibir razones es la característica de la identidad humana (§5.2.3). En efecto, la distancia reflexiva que podemos tomar con respecto a nuestros deseos, creencias, incentivos, inclinaciones, etc., nos obliga a explicar en qué sentido estas propuestas pueden ser una razón de nuestra acción. Luego la capacidad de dar y recibir razones es intrínseca a nuestra condición de humanidad. Esta aclaración es importante en este punto del argumento porque empata con la noción de reino de los fines. La noción está comprometida con la idea de que los miembros de este reino son seres con la capacidad de dar y recibir razones, esto es, son los seres humanos (racionales, por definición) quienes forman el reino de los fines y quienes tienen la capacidad de asentir universalmente a ley moral.

La idea misma de reino de los fines no puede entenderse al margen de la identidad humana. Al trazar el puente entre el reino de los fines y la identidad humana se da el salto al terreno moral. Atrás había asumido, junto con Korsgaard, la idea de que un agente que valora su identidad humana necesariamente ha de reconocer el valor de la identidad humana de otros seres humanos. En otras palabras, la idea misma de identidad humana no puede concebirse sin extenderse a todos los seres humanos racionales. Este supuesto puede explicarse y justificarse a través de las razones públicas: si nuestra identidad humana se caracteriza por la autoconsciencia y, por tanto, por la capacidad de dar y recibir razones públicas, resulta que no puedo concebir ni valorar mi identidad humana sin otorgar esta misma capacidad a otros seres humanos. No tendría sentido dar ni recibir razones si no hay otro ser humano que recíprocamente reciba y aporte razones en el espacio público. Luego, aquello que valoro de

mi identidad humana ha de ser valorado también en otros seres humanos racionales (seres que dan y reciben razones). Mi identidad humana no puede concebirse al margen de la identidad humana de los otros.

De esta manera se explica mejor el sentido en que la identidad humana es también una identidad moral. Si valoro la humanidad de otros seres humanos, los estoy considerando como agentes que actúan de acuerdo con razones y, por lo tanto, como autores de sus propias acciones. Y esto significa también que no puedo considerarlos solo como medios. Sus razones han de tener peso normativo en el espacio público de razones, compromisos normativos morales. Aquel que pueda actuar con base en razones no puede ser tratado como simple medio. En virtud de su capacidad racional quienes pueden actuar con base en razones tienen un valor intrínseco, tienen dignidad. Y es así como se explica el paso del asentimiento universal (imperativo categórico) a la ley moral. La ley moral se fundamenta en la identidad humana que todos los seres racionales tienen en virtud de la condición humana compartida.

c. Integridad y unidad

El recorrido por la noción de identidad práctica y por la distinción entre razones públicas y privadas es relevante para explicar finalmente cómo el modelo constitucional integra o unifica al agente.

Las razones que tiene un agente para actuar se fundan en las diferentes identidades prácticas con las que él esté comprometido. Estas identidades proponen un curso de acción, proponen razones para actuar. Y, según el modelo constitucional, la función de la identidad moral es legislar sobre estas propuestas. Sin embargo, como he mostrado, las propuestas de las identidades no-morales solo pueden ser validadas si se adecúan a los parámetros del imperativo categórico. Lo anterior significa que el agente integra sus distintas identidades siempre que todas ellas estén de acuerdo con su identidad moral. Es en este sentido que "la ley moral es la ley de la auto-constitución [del agente]" (Korsgaard, 2009, p. 214).

Lo anterior muestra que la relación entre las identidades prácticas no-morales y la identidad moral es indispensable para entender la manera en la que el agente está integrado. Como

dije, las identidades no morales generan razones para la acción. Si esta razón es privada y no cumple el requisito del imperativo categórico, esto es, si no tiene pretensiones de validez universal, entonces se está poniendo en peligro la integridad del agente, pues una razón privada es una forma de fragmentar la agencia, es una forma en la que el yo presente se separa o desintegra con respecto a sus otros yoes posibles-futuros. Una razón pública que, por el contrario, se define como sometida a los estándares normativos públicos hace posible la agencia unificada[75], pues el yo de ahora está de acuerdo con cualquier yo-posible, y por tanto, el agente se mantiene como una totalidad unificada y estable en el tiempo. "La fuerza normativa compartida es el adhesivo [*glue*] que mantiene al agente reunido" (Korsgaard, 2009, p. 204).

La relación entre razones públicas e integridad es de este modo directa. La conclusión a la que llega Korsgaard es que las razones públicas, que son razones que han de ser asentidas universalmente, son las únicas que unifican al agente. En otras palabras, la unificación solo es posible cuando el agente cuenta con razones válidas ajustadas al imperativo categórico. "Para tener una voluntad unificada, debes desear de acuerdo con una ley universal. De otra manera, tú eres solo un manojo de impulsos, y no un agente después de todo" (Korsgaard, 2009, p. 213).

Y dado que, además, el asentimiento universal presupone el reconocimiento de la identidad humana y de la capacidad para recibir y dar razones, entonces resulta que tener una razón pública exige comprometernos de entrada con la ley moral, pues actuar según una razón pública presupone asumir la perspectiva de una comunidad de seres racionales en la que todos legislan de acuerdo con el imperativo categórico y en la que comparten su condición de humanidad (específicamente, comparten la capacidad de dar y recibir razones). Con este análisis se llega a una conclusión de cuño kantiano: actuar de acuerdo a la ley moral es una condición para ser un agente unificado. En términos más abstractos, la unificación o la integridad del agente son posibles gracias a la moralidad.

---

[75] "La acción es simplemente interacción con el yo. Si esto es así, entonces respetar a la humanidad en la propia persona, y el consecuente tratamiento de las propias razones como consideraciones con calidad normativa y pública, son las condiciones que hacen posible la agencia unificada" (Korsgaard, 2009, p. 204).

> [L]a ley moral es la ley de la constitución unificada, la ley de la persona de quien realmente puede decirse que legisla por sí misma porque ella es la persona que realmente tiene un yo [self]. Esta es la ley de una auto-constitución exitosa. (Korsgaard, 2009, p. 206)

Lo que unifica o integra al agente es la identidad moral-humana. Es ella la que, siguiendo el modelo constitucional, legisla con respecto a las otras identidades no-morales. Las identidades no-morales y la identidad moral se integran o se unifican cuando las razones de las identidades no-morales pueden ser asentidas universalmente de acuerdo con el imperativo categórico y de acuerdo con la ley moral. Cuando esta integración se da, el agente actúa como un todo unificado.

Y además es a través del ejercicio de la identidad moral que el agente logra auto-constituirse, pues cuando decide según razones públicas (razones universales para una comunidad de seres racionales) él está ejerciendo su capacidad de ser la causa de sus acciones y, por lo tanto, está constituyéndose como autor de ellas. Como se ve, Korsgaard, a través del modelo constitucional, logra articular la idea de que un agente auto-constituido es un agente integrado.

La interpretación llevada a cabo por Korsgaard de la integridad como unidad respeta la etimología del concepto: "Etimológicamente, la integridad es unidad [*oneness*], la integración es lo que hace que algo sea uno. […] En el sentido metafísico, eso es lo que significa tener integridad" (Korsgaard, 1996, p. 102). Lo interesante es que con la idea de identidad práctica se puede mantener el sentido de unidad pero ahora en un plano práctico: "nosotros usamos el término [integridad] para alguien que vive de acuerdo con sus propios estándares. Y esto es porque pensamos que vivir de acuerdo con ellos es lo que hace de él *uno*, y así lo que hace de él una persona como tal" (*ibid.*; énfasis añadido). Esta afirmación significa que el agente, cuando guía su acción por razones que son válidas universalmente, ha logrado unificar o integrar sus identidades no-morales con la identidad moral.

5.3 Integridad como auto-constitución

¿Cómo responde Korsgaard al problema de la integridad personal? El primer movimiento de la autora es mostrar que la identidad moral de una persona es condición de posibilidad de ser-agente constituido. El segundo movimiento consiste en considerar que la identidad moral es aquella que unifica al resto de identidades prácticas en torno a la identidad moral[76]. Este es el sentido en el que un agente se auto-constituye como un todo integrado. La identidad humana-moral es entonces el punto arquemidio en donde está ubicada la agencia. En un estilo marcadamente kantiano, lo que en últimas integra al agente, lo que lo constituye, no es otra cosa que su identidad moral. La agencia no puede pensarse sin la idea de integridad, y la integridad no es posible sin la identidad moral.

La articulación que traza Korsgaard entre el modelo constitucional y la noción de integridad en términos de unidad práctica parece responder al problema de la integridad personal. Korsgaard estaría diciendo, contra Williams, que la moralidad kantiana no representa una amenaza a la integridad personal; todo lo contrario, la identidad moral es condición de la unificación o integración de la persona.

Este resultado tan opuesto a Williams, como ya lo he mencionado, supone una interpretación muy distinta de la noción de integridad. Como he mostrado en el capítulo 4, Williams está comprometido con la integridad entendida en términos del carácter y los proyectos fundamentales de una persona. Poner en peligro alguno de estos elementos supone una

[76] El argumento que presenta Korsgaard en *Sources of Normativity* (1996) va en la misma dirección que la idea de *Self-Constitution* (2009), solo que en el primer texto la autora no se centra en la noción de agencia, sino en la de humanidad. La autora argumenta que valorar cualquier cosa depende de valorarnos a nosotros mismos como seres humanos (Rivera, 2007, p. 79n21). La humanidad es la fuente última de la normatividad. Y es a partir de la idea de humanidad que se justifica la adopción del imperativo categórico como principio de la racionalidad práctica, pues —según ella— es este principio el único compatible con el respeto a la humanidad tanto en mi persona, como en la de cualquier otro. Rechazar la autoridad del imperativo es negar el valor de la humanidad, y por tanto, supone perder nuestra capacidad para valorar en general. "La identidad moral ejerce un tipo de gobierno sobre las otras identidades. Las concepciones prácticas de tu identidad que sean fundamentalmente inconsistentes con el valor de la humanidad deben ser rechazadas" (Korsgaard, 1996, p. 129s). Independientemente de la diferencia argumentativa, Korsgaard apunta a la misma intuición en ambos textos: la moralidad (identidad moral) tiene primacía normativa con respecto al resto de identidades prácticas.

pérdida de integridad. Por su parte, Korsgaard está entendiendo la integridad en términos de unidad*:* el factor relevante de la integridad no es preservar el carácter, sino la unidad del agente.

La diferencia entre las dos aproximaciones, sin embargo, no se reduce a un choque de intuiciones. No solo se trata de que Williams privilegie cierto campo semántico del concepto de integridad, mientras que Korsgaard privilegie otro aspecto. Tal y como interpreto la posición de Korsgaard, su propuesta además explica por qué es mejor entender la integridad en términos de unidad y no en términos de carácter: ser agente unificado, para ella, es una condición más básica que la condición de tener un carácter —en los términos de Williams. En otras palabras, tener un carácter o tener proyectos fundamentales es algo que ya presupone ser-agente. En este sentido, la propuesta de Korsgaard está un paso atrás de la propuesta de Williams. El punto de partida de Williams es que ya somos agentes constituidos con carácter y con proyectos fundamentales, y su idea es que en ocasiones las exigencias imparciales e impersonales de la moralidad pueden amenazar la integridad de este agente ya constituido. Korsgaard, por su parte, responde que la condición de posibilidad para ser agente es que haya una identidad moral básica que unifique el resto de identidades prácticas.

La propuesta de Korsgaard apela a un argumento trascendental: la condición de posibilidad de ser agente es la auto-constitución. La auto-constitución unifica o integra al agente, y además esta integración o unificación solo es posible en virtud de la identidad moral. Y esta identidad moral es más básica que cualquier otra identidad práctica. Por lo tanto, la identidad moral tiene primacía normativa respecto de las demás identidades prácticas. Al aceptar el argumento de Korsgaard, también se está aceptando el peso normativo de la identidad moral. Esta es tan fundamental y básica para la agencia, que ir en contra de ella —como Williams parece sugerir— supondría la desintegración del agente en tanto que agente.

En su propuesta, Korsgaard no está negando que se puedan presentar conflictos entre la identidad moral y las otras identidades prácticas. Esto es algo en lo que ella estaría de acuerdo con Williams. Su punto es más bien que la identidad moral, al ser más básica en la constitución de la agencia, tiene primacía normativa. En este sentido, un conflicto con la

identidad moral habrá de resolverse a favor de la identidad moral, pues esta es constitutiva de la agencia, mientras que la identidad práctica no-moral es secundaria.

6. Crítica a la primacía normativa de la moralidad: sentido y necesidad práctica

6.1 Crítica al modelo constitucional

6.1.1 Identidades prácticas y asimetrías

6.1.2 La tesis de la indispensabilidad de los principios universales

6.2 ¿Primacía normativa de la moralidad?

6.2.1 Prioridad constitutiva de la moralidad

a. Schmagencia [*Schmagency*] y el juego de la agencia

b. 'Razón para estar vivo' y moralidad

c. ¿Primacía normativa de la moralidad?

d. Reivindicación de la crítica de Williams: moralidad y sentido

6.2.2. Tesis de la prevalencia

a. La distinción entre ética y moralidad

b. El concepto de importancia

c. Prioridad deliberativa

d. Necesidad práctica

e. Obligación moral y no-moral

La solución de Korsgaard al problema de la integridad consiste en señalar que la integridad entendida en términos de unidad y auto-constitución es más básica que la integridad entendida en términos de identidad. De esta manera, diría Korsgaard, un agente con carácter y con proyectos —como el que tiene en mente Williams— es un agente que previamente ha de estar unificado como un todo integrado. Y como vimos, esta integración compromete al agente con el imperativo categórico y con la ley moral. De manera que el agente de Williams es ya un agente moral (con todo el peso kantiano que esto implica). Y lo que es aún más importante, este compromiso moral es condición de posibilidad de su ser-agente. Lo anterior daría para pensar que la crítica de Williams a las exigencias incondicionales de la moralidad empieza a perder fuerza, pues en un

conflicto entre la moralidad y la integridad, no podemos simplemente renunciar a la moralidad porque esto implicaría perder la condición de agencia. A la idea de Williams según la cual las exigencias incondicionales de la moralidad pueden amenazar la integridad de la persona, Korsgaard responde que la noción importante de integridad no es aquella que preserva el carácter de la persona, sino aquella que unifica al agente en un todo integrado, y por tanto su postura no amenaza la integridad de la persona. Así que en contra de lo que piensa Williams, Korsgaard logra armonizar la incondicionalidad de la moralidad con la integridad del agente.

El rigorismo clásico de Kant estaba comprometido con la tesis de la primacía normativa del deber (§1.2, §1.3 y capítulo 2) y con la concepción del yo dividido (capítulo 3). En mi interpretación, el rigorismo kantiano conducía al problema de la integridad personal (capítulo 4). Korsgaard, al parecer, logra responder a este problema (capítulo 5). Su propuesta implica rechazar el modelo de la agencia del yo dividido en favor del modelo constitucional. Con este modelo ella espera resolver el problema de la integridad personal sin renunciar a la primacía normativa de la moralidad.

Korsgaard se separa de la versión clásica del rigorismo, rechaza la concepción del yo dividido pero mantiene, en un sentido muy fuerte, su núcleo normativo. En mi interpretación, entonces, esta autora sigue comprometida con lo que llamo la tesis de la *primacía normativa de la moralidad*. Esta tesis, como ya he sugerido antes, sostiene que la moralidad es el valor normativo más importante o tiene prevalencia sobre cualquier otro criterio práctico. La consecuencia de la primacía normativa de la moralidad, al igual que sucede en el caso del rigorismo, es que la deliberación del agente deberá estar guiada por la normatividad moral.

La diferencia entre el rigorismo moral de Kant y la posición de Korsgaard es que esta última no niega la posibilidad de que existan otras fuentes de valor distintas a la moralidad. Si bien ambos autores sostendrían que la moralidad es un valor absoluto e incondicional, Kant además sostiene que únicamente la moralidad tiene valor genuino (por sí mismo), mientras que el resto de valores serían instrumentales (valiosos como medio para otra cosa) (*F* 4:428). Korsgaard, por su parte, acepta que además de la moralidad también existen otros valores no morales que pueden ser incondicionales. Ella usa la expresión "identidad práctia incondicional" para referirse a esta idea. Su posición representa una versión matizada del rigorismo moral kantiano.

> [E]l argumento no muestra que todas las obligaciones sean morales, ni que las obligaciones morales siempre tengan prioridad sobre las demás. […] La identidad moral no hace zozobrar otras formas de identidad, pues nadie es simplemente moral y nada más. Tiene razón Bernard Williams cuando afirma que si la moralidad nos exigiera eso sería incoherente. Sin embargo, no sería correcto concluir que, por lo tanto, ni la obligación moral, por un lado, ni nuestras demás obligaciones, por el otro, pueden ser incondicionales. […] Las obligaciones en conflicto pueden ser incondicionales la una y la otra; no es más que una de las maneras en que la vida humana es difícil. (Korsgaard, 1996, p. 125s)

Con estas afirmaciones, insisto, Korsgaard está tomando distancia con respecto al rigorismo clásico kantiano. En mi lectura, la diferencia más sustancial entre su propuesta y la de Kant está en sus distintas posiciones con respecto a los conflictos prácticos. Mientras que Kant no acepta la existencia de dilemas morales ni de conclictos prácticos (*MC* 224), Korsgaard, en cambio, admite la posibilidad de que la moralidad entre en conflicto con las obligaciones de otras identidades prácticas.

> [T]enemos que distinguir dos clases de conflicto. Una persona puede tener una identidad práctica que sea de suyo contradictoria con el valor de la humanidad, por ejemplo la identidad de un asesino; o bien tener una identidad práctica que no sea por naturaleza contraria al valor moral, pero que en una u otra situación entre en conflicto con él. Creo que el curso de reflexión que he tratado de describir excluye la primera clase de identidad y los conflictos que genera. En la medida en que la importancia de tener una identidad práctica proviene del valor de la humanidad, no tiene sentido que uno se identifique de maneras que sean inconsistentes con el valor de la humanidad. La segunda clase de conflicto, empero, no puede excluirse de esta manera; todavía es posible que las exigencias concretas de la moralidad entren en conflicto con las [exigencias] de alguna forma más contingente de identidad. (Korsgaard, 1996, p. 126)

Con este movimiento, Korsgaard distingue dos tipos de conflicto. Aquellos que denomino conflictos anti-morales porque suponen que una de las partes es contradictoria con el valor moral

de la humanidad. Los casos del botánico y del químico serían ejemplos de este tipo de conflicto. Para Korsgaard, los conflictos anti-morales no son conflictos reales, pues no se puede renunciar al valor de la humanidad y, por tanto, el conflicto se ha de resolver siempre a favor de la moralidad que se corresponden con el valor de la humanidad. Este es un claro punto de contacto con el rigorismo clásico. Un segundo tipo de conflictos son lo que en la literatura se denominan conflictos prácticos (Bauman & Beltzer, 2004). A diferencia de un dilema moral que enfrenta dos valores morales, en los conflictos prácticos la tensión se da entre valores morales y no-morales (p.e. el conflicto de Gauguin entre cuidar a su familia y perseguir su proyecto de ser pintor). Korsgaard acepta este segundo tipo de casos.

Korsgaard reconoce entonces que el conflicto práctico es posible, solo que, según su propuesta, la posibilidad del conflicto "pierde su aguijón" (Korsgaard, 1996, p. 126). La autora cree que al reconocer la posibilidad del conflicto, su posición se desmarca de las consecuencias rigoristas del modelo kantiano. No creo, sin embargo, que esto sea cierto. Reconocer que los valores no-morales pueden ser incondicionales no significa que la moralidad no tenga prioridad sobre los valores no-morales. En un conflicto entre valores incondicionales todavía está abierta la pregunta práctica acerca de cuál valor hemos de adoptar. Y es en este sentido como la moralidad todavía tiene una prioridad normativa: esta es constitutiva de nuestra condición de agentes, lo cual supone que tenemos siempre una razón para preferir la moralidad por encima de los valores no-morales.

En este capítulo, mi objetivo es cuestionar la tesis de la prioridad normativa de la moralidad de Korsgaard. Mi argumento está articulado en dos partes. La primera es una crítica al modelo constitucional. Siguiendo a Josep Corbí, muestro que este modelo de la agencia todavía sigue comprometido con la concepción del yo dividido (§6.1). La segunda parte es una crítica a la tesis de la primacía normativa de la moralidad (§6.2). Distingo entre dos versiones de esta tesis: la versión fuerte o la tesis de la primacía constitutiva de la moralidad (§6.2.1.) y la versión débil o la tesis de la prevalencia (§6.2.2). Como espero mostrar, ambas son falsas, y por tanto, en contra de Korsgaard, la prioridad normativa de la moralidad no puede sostenerse.

6.1 Crítica al modelo constitucional

El Modelo Constitucional propuesto por Korsgaard conlleva ciertas consecuencias conceptuales y supone ciertos compromisos con algunas tesis que, vistas con atención, resultan problemáticas. Una de las críticas a este modelo es desarrollada por Josep Corbí, su objeción radica en señalar que este modelo todavía está comprometido con la concepción del yo dividido (Corbí, 2012; 2015).

La estrategia de Corbí consiste en mostrar que el Modelo Constitucional comporta una inconsistencia interna. Su crítica, como la entiendo, se estructura en dos momentos. En primer lugar, la distinción entre identidad moral e identidad no-moral genera unas "asimetrías" que solo pueden "entenderse desde la concepción escindida del sujeto" (Corbí, 2015, p. 39). En segundo lugar, se intenta mostrar que la tesis de la indispensabilidad de los principios universales —el sujeto solo se constituye por principios morales o de lo contrario sería un amasijo de impulsos— solo es plausible si, nuevamente, nos comprometemos con la concepción del yo dividido.

6.1.1 Identidades prácticas y asimetrías

El concepto de identidad práctica con el cual Korsgaard está comprometida es, según ella, esencial para poder explicar la unidad del agente. Hemos visto que la autora también reconoce que existe una distinción entre la identidad moral y las identidades no-morales y que la identidad moral tiene un privilegio sobre el resto de identidades prácticas.

> Casi todo el tiempo, nuestras razones para la acción se originan en nuestras identidades más contingentes y locales; sin embargo, parte de la fuerza normativa de estas razones se origina en el valor que nos conferimos en tanto seres humanos que necesitan tales identidades. De esta manera, todo valor depende del valor de la humanidad; otras formas de identidad práctica importan, en parte, porque la humanidad las necesita. La identidad moral y las obligaciones que trae consigo son, en consecuencia, ineludibles y omnipresentes. Después de todo, no toda forma de identidad práctica es contingente o relativa; la identidad moral es necesaria. (Korsgaard, 2000a, p. 154)

Para Korsgaard, como vimos, la identidad moral es esencial. Es esta la que dota de cohesión y unidad al agente. Sin ella el sujeto "se descompondría en un amasijo de deseos inconexos"

(Korsgaard, 2009, p. 204). Sin embargo, la distinción entre identidades necesarias o esenciales, por un lado, y contingentes, por otra, es muy problemática. ¿Cómo puede una persona distanciarse de sí misma, es decir, apartarse de cualquiera de sus identidades contingentes? ¿Cómo puede este mismo movimiento no ser posible con respecto a la identidad moral? (*cf.* Corbí, 2012, p. 130). Para Korsgaard, la distancia que uno puede tomar con respecto a las identidades prácticas no-morales significa que estas son contingentes en la constitución de nuestras vidas (§5.2.3.a), pues, por un lado, no tienen suficiente fuerza normativa y, por otro, no afectan sustancialmente o no afectan en lo absoluto la posibilidad de ser-agente. En cambio, no podemos distanciarnos de la identidad moral, sus exigencias normativas son totalmente vinculantes y, en caso de no cumplirlas, perdemos nuestra condición de agentes.

El concepto de 'tomar distancia' o 'distancia reflexiva' es relacional. Al menos en términos geométricos, la distancia supone siempre dos puntos con respecto a los cuales se mide una magnitud que los separa. Esta analogía se puede trasladar para el caso que nos ocupa. El punto que tomamos como referencia para medir la distancia es la identidad real del agente[77], su identidad moral. Y es desde la perspectiva de este sujeto real-moral que es posible separarse respecto de las identidades contingentes. Guardando esta analogía geométrica, la identidad moral es el punto arquemidio en donde está ubicada la agencia. Lo que está más allá de este centro es contingente con respecto a lo que constituye realmente al agente.

Evidentemente, como señala Corbí, aquí se presenta una primera asimetría entre la identidad moral y las identidades no-morales. Se trata de una asimetría que se concibe en términos de necesidad y contingencia: en la constitución de la agencia, la identidad moral es necesaria; las demás son eludibles y contingentes.

Una segunda asimetría entre identidad moral y no-moral puede encontrarse, según Corbí, examinando al papel de la identidad moral. Cuando describí la psicología no empirista que subyace al modelo constitucional (§5.2.2), señalé que las identidades prácticas contingentes son vistas como

[77] Esta analogía geométrica evoca la teoría del yo real: existe un yo real que es el agente, y que constrasta con otra parte que es secundaria o dependiente de este yo real (§3.4).

propuestas que se presentan para ser sometidas a la reflexión y evaluación de la identidad moral. Es desde este tribunal como se decide si la identidad contingente es adecuada o no lo es.

> Incluso aunque la identidad práctica [contingente] pueda ser explícitamente reconocida como muy significativa en la vida del agente, el núcleo de su identidad es todavía su identidad moral, en la medida en que es indispensable y autosuficiente para evaluar las propuestas que puedan provenir de las otras partes del sí mismo. (Corbí, 2012, p. 131)

Sin embargo, el compromiso con la psicología no empirista implica que la razón tiene una autoridad que no tienen las demás identidades prácticas, es ella la que asiente o rechaza las propuestas de las identidades prácticas particulares. La identidad moral del sujeto se convierte así en la expresión de la voluntad real de una persona. La voluntad del sujeto es, y solo puede ser, una voluntad moral. Este rasgo, a todas luces kantiano, lleva a Korsgaard a concluir que el agente moral puede llegar a escapar de la cadena causal, de manera que solo se constituye realmente como agente gracias a su identidad moral:

> Hacer que algo sea objeto de la voluntad no es sólo ser una causa, ni siquiera es permitir que algún impulso actúe en mí como causa, sino, por así decirlo, tomar las riendas conscientemente, y hacerme a mí mismo la causa de lo que hago. Y si voy a constituirme como la causa de una acción, entonces debo ser capaz de distinguir entre que yo cause la acción y algún deseo o impulso que esté «en mí» y cause que mi cuerpo actúe. Debo ser capaz de verme a mí mismo como algo distinto de cualquiera de mis impulsos y motivos particulares de primer orden, tal como lo exige la perspectiva reflexiva en cualquier caso. Mínimamente, entonces, no soy la simple ubicación de un deseo causalmente efectivo, sino que soy, más bien, el agente que actúa conforme al deseo. Por lo anterior, si asiento a actuar de cierta forma ahora, debo al mismo tiempo asentir a actuar del mismo modo en todas las ocasiones que sean similares de manera pertinente. (Korsgaard, 2000a, p. 279ss; énfasis en el original)

Para que la razón pueda ejercer su función y pueda constituirse en una autoridad debe poder separarse o tomar distancia, en la medida de lo posible, de las identidades no-morales. De lo contrario no podría ejercer su papel imparcialmente. Es por esta razón que Korsgaard aboga por

la idea de que la deliberación práctica sea formal, es decir, una en la que lo relevante sea el procedimiento a través del cual se lleva a cabo la deliberación moral.

Al igual que sucede en la primera asimetría, el agente tiene una identidad moral que toma distancia con respecto a las identidades no-morales. Estas últimas se convierten en objeto de reflexión de la identidad moral (Corbí, 2015, p. 44). Evidentemente, si estas identidades son objeto de la reflexión, no pueden hacer parte del entramado de la identidad moral:

> De otro modo, deberíamos aceptar que la identidad moral desde la que se pondera cada incentivo o propuesta está parcialmente moldeada por algunas de las identidades no-morales del sujeto, de manera que la identidad moral podría variar de un individuo a otro según cuáles fuesen las identidades no-morales que estuviesen involucradas en su procedimiento evaluativo. (Corbí, 2015, p. 44).

Para alcanzar la imparcialidad, la identidad moral debe permanecer libre de cualquier interferencia de las otras identidades prácticas.[78] El agente constituido moralmente ha de guardar distancia con respecto a sus otras identidades. ¿No es esta una forma de trazar una división en el sujeto, o en otras palabras, no es esto una reproducción de la concepción del yo dividido?

Incluso si, con el propósito de evitar la escisión del agente , Korsgaard admitiera la posibilidad de que las identidades no morales hicieran parte de la identidad moral, entonces "se difuminarían […] los límites entre la identidad moral de un sujeto y sus otras identidades prácticas" (Corbí, 2015, p. 44). La única forma de mantener el privilegio y la autoridad de la identidad moral es aceptando que esta es independiente de otras identidades prácticas, es decir, una identidad moral pura.

Existen entonces dos asimetrías entre las identidades morales y las no-morales. La primera se

[78] Si la identidad moral no es pura, entonces existiría una divergencia moral entre las personas: cada una evaluaría según una identidad moral permeada por identidades prácticas muy distintas. Aun si todas las personas siguieran el mismo criterio formal de deliberación moral —que en este caso se corresponde al imperativo categórico— el hecho de que la identidad moral esté contaminada por la contingencia afecta el resultado de esta deliberación. Por tanto, no podría alcanzarse un acuerdo universal y, en ese sentido, la identidad moral resulta ser una identidad particular y hasta idiosincrática.

expresa en el hecho de que la identidad moral es esencial para la constitución del agente, mientras que las otras identidades prácticas son contingentes y prescindibles. La segunda asimetría consiste en que la identidad moral tiene una autoridad que no tienen las otras identidades. Esta autoridad solo puede garantizarse bajo la pretensión de un tribunal imparcial y puro que excluye la posibilidad de que la identidad moral atienda enteramente a las necesidades de las identidades no morales. Ambas asimetrías dejan muy clara la relación que se establece entre las identidades del sujeto. De acá que se pueda concluir que el modelo constitucional está comprometido con la concepción del yo dividido "dado el modo en que el yo moral se sitúa frente al resto de las partes del sujeto" (Corbí, 2012, p. 45). La diferencia respecto del modelo del combate es que esta división ya no se expresa en términos razón-inclinación, sino en términos de la división entre la identidad moral y las identidades no-morales.

6.1.2 La tesis de la indispensabilidad de los principios universales

Además de las asimetrías que he presentado, Corbí sostiene que también existe otra razón para demostrar que el modelo constitucional de Korsgaard está comprometido con la concepción del yo dividido. Esta razón tiene que ver con la tesis de la *indispensabilidad de los principios universales*. Como ya he señalado, en la propuesta de Korsgaard, los principios universales tienen la función de unificar al agente, pues, según ella, el sujeto solo puede constituirse adoptando principios morales. Son ellos lo que lo salvan de convertirse en un amasijo de impulsos inconexos. Lo paradójico, como espero mostrar, es justamente que este intento por unificar al sujeto conlleva a su desintegración. Veamos.

Para constituirse como sujeto, dice Korsgaard, es necesario que los impulsos estén hasta cierto punto unificados (*cf.* 2009, p. 27). De hecho, ella parece estar planteando un dilema: o adopto principios universales o me convierto en un amasijo de impulsos. Si fuéramos un amasijo de impulsos la agencia es imposible, la única forma de constituirse como agente es, entonces, a través de la adopción de principios universales. La lectura que hace Corbí de Korsgaard, y que yo suscribo es, entonces, aquella según la cual la unificación del sujeto solo es posible "a través de la fidelidad a los principios o leyes que la voluntad asuma como propios" (Corbí, 2015, p. 45).

> El trabajo de la deliberación práctica es la reunificación, la reconstitución: y la función de los principios que gobiernan la deliberación —los principios de la razón práctica— es la unificación del yo. La función de la razón práctica es unificarnos como agentes que podamos ser autores de nuestras acciones. (Korsgaard, 2009, p. 126)

De cualquier forma, esta unificación no tiene lugar sin más, ella requiere un esfuerzo por parte del agente. Korsgaard también reconoce este mismo punto:

> Hay trabajo y esfuerzo –una especie de lucha– involucrada en la vida moral, y los que tienen éxito los denominamos "racionales" o "buenos"; pero no es la lucha por ser racional o bueno. Es, más bien, la lucha por la integridad, la lucha por la unidad psíquica, la lucha por ser un agente unificado a pesar de la complejidad psíquica. (Korsgaard, 2009, p. 7)

Dado lo anterior, entonces los elementos no-racionales deben ser sensibles o permeables a las leyes de la moralidad (Corbí, 2012, p. 132). La razón, a través de este esfuerzo, debe poder afectar estas partes no racionales. Si la parte no-racional no fuera permeable a la razón, no sería posible la unidad en el agente porque habría en el ser humano dos fuerzas heterogéneas e inconmensurables

La permeabilidad de la parte no-racional, sin embargo, no es absoluta. Es un hecho que lo no-racional no siempre se adecúa a la razón. Este hecho denota cierta resistencia en contra de la razón. Desde el punto de vista de Kant, esta resistencia debe poder ser vencida gracias a la fuerza o al esfuerzo de la voluntad, ella debe poder ejercer un influjo causal sobre nuestra naturaleza fenoménica (*F* 4:408, 4:410-11). Tanto Kant como Korsgaard confían en que la voluntad llegue a determinar a las inclinaciones, y además que las partes no-racionales sean permeables a este poder.

Sin embargo, esta resistencia no es un residuo que pueda simplemente dejarse de lado como lo que sobró de la unificación racional del agente. Al contrario, la resistencia y el residuo indican que la vida práctica de la persona no se agota en lo que pueda determinar la razón (Corbí, 2012, p 132). Es justamente el residuo lo que prueba que la unificación a partir de principios morales universales no se produce. En otras palabras, la resistencia generada por la parte no-racional es una muestra de que la integración del agente no es completa, a menos, claro, que consideremos que lo único

que cuenta como agencia sea la parte racional.

Una vez se logra la unificación, en ¿qué condición queda el residuo que no es permeable a la razón? ¿Acaso el residuo debe ser considerado extraño y ajeno respecto a la vida práctica de la persona? Si así fuera, es evidente que la unificación que pretende el modelo constitucional está introduciendo una división entre aquello que pertenece a la agencia y que es un todo integrado, y aquello que es un residuo que no es permeable a la razón. En otras palabras, en la unificación a través de principios universales se excluyen completamente otros elementos no-racionales. ¿Pero se puede reducir la vida práctica del ser humano a la agencia racional? Eludir o pretender suprimir esas identidades prácticas que no se adecúan enteramente a la razón es una forma de racionalizar excesivamente la vida humana.

> [L]a presencia de esta resistencia debería ser considerada como una invitación para revisar la propia ley y, consecuentemente, esa unificación no siempre será el resultado de la capacidad de la voluntad para imponer cierta ley, sino que algunas veces podría requerir cierta sensibilidad a las otras partes de uno mismo y en últimas a la situación particular que uno podría enfrentar. (Corbí, 2012, p. 132)

Korsgaard estaría de acuerdo en aceptar que las partes no racionales de un sujeto de alguna manera sean reconocidas por el tribunal de la identidad moral. Así que las identidades prácticas no racionales han de estar integradas a la constitución del sujeto. La forma como la autora intenta llevar a cabo este paso es introduciendo la distinción ya mencionada entre razones públicas y razones privadas.

Las razones públicas, como vimos, son razones válidas objetivamente en virtud de que están en concordancia con el imperativo categórico. Esto explica el que puedan ser razones a las que asientan todos los seres racionales, incluyendo, por supuesto, al agente mismo. En cambio, una razón privada no necesariamente está en concordancia con el imperativo y, por lo tanto, no es necesario que sea asentida universalmente. Cuando una razón privada no puede ser universalizable, ni siquiera uno mismo podría estar de acuerdo con ella; pero si una razón privada puede ser pensada como una ley universal, entonces también puede ser integrada junto a la propia identidad moral. Las razones privadas que surgen de las identidades no-morales son sometidas al

tribunal de la razón, y para que sean aprobadas deben poder contar como razones públicas, es decir, deben de alguna manera adecuarse a los criterios normativos de la identidad moral. Si las identidades prácticas no contaran como razones públicas, simplemente no pasarían el arbitraje de la razón, y, por lo tanto, el agente no podría asentir a ellas. La identidad moral no podría validar unas exigencias que provengan de razones privadas.

A partir de este análisis resulta que la permeabilidad es central en la propuesta de Korsgaard. Si el modelo constitucional funciona adecuadamente, basta el dictamen de la razón para que las identidades no-morales obedezcan y se repliguen a los dictados de la identidad moral. Sin embargo, el hecho de que una de las partes del yo se resista a las decisiones de la voluntad es muestra de que la razón no permea completamente a la totalidad del agente. La resistencia misma prueba que "la voz de alguna de las partes del yo no ha sido debidamente ponderada" (Corbí, 2015, p. 46). En otras palabras, puede que haya razones privadas que, aun cuando no sean razones públicas, sí tienen un lugar importante en la conformación de nuestra identidad.

Para Corbí, la resistencia que ofrecen las identidades no-morales a la voluntad es una prueba de que la integración del agente no se agota en el procedimiento de la razón. Si la razón ha emitido el veredicto, se esperaría, gracias a la permeabilidad, que las partes no-morales obedecieran y acallaran sus propuestas. Pero no es esto lo que normalmente sucede. La voz de las partes no-morales aún se escucha, a pesar de los dictados morales. Esta es justamente la descripción de la experiencia de conflicto práctico. Si esperamos que la 'fuerza de la voluntad' aplaque esta resistencia, o si simplemente desatendemos a la voz no-moral que expresa su resistencia, el sujeto estaría, en cualquier caso, "condenado a una forma de fragmentación" (Corbí, 2015, p. 46).

Por lo visto, el intento de Korsgaard por unificar al sujeto alrededor de la identidad moral empieza a debilitarse. Según Corbí, la propuesta de Korsgaard enfrenta un problema al tratar las identidades no morales como meras propuestas. Se supone que estas han de estar sometidas a los criterios morales y que son permeables a leyes morales. Sin embargo, la resistencia de las partes no-racionales demuestra que después de todo ellas no son tan permeables. Para Corbí, el fracaso de la voluntad moral en su esfuerzo por permear todas las partes del yo puede deberse a la "inadecuación de la ley que la voluntad se ha dado a sí misma y, en definitiva, del procedimiento

deliberativo que ha conducido a su elección" (Corbí, 2015, p. 46).

La resistencia de las partes no-morales contra las pretensiones de la voluntad da cuenta, entonces, del hecho de que la tesis de la indispensabilidad de las leyes universales puede ser cuestionada. Korsgaard cree que si no adoptamos principios universales nos convertiríamos en un amasijo de impulsos. Sin embargo, adoptar los principios universales no necesariamente da lugar a la unificación del agente, pues con frecuencia persisten elementos que no son permeados por estos principios. Despreciar la voz de las identidades no-morales es una manera de desconocer una parte de uno mismo, de silenciar forzadamente algo que también forma parte de uno mismo. Este desconocimiento es, en sí mismo,una forma de fragmentación. Es esta la razón por la cual la tesis de la indispensabilidad compromete al modelo constitucional con la concepción del yo divido.

Creo que el análisis que lleva a cabo Corbí conduce a un resultado muy negativo para el modelo constitucional. Si la razón únicamente evaluara las propuestas no-morales en términos formales y bajo criterios universales, entonces, frente a la resistencia de las partes, nos vemos abocados a mantener una imagen del sujeto como fragmentado. Si la unificación solo es posible a través de principios universales, pero estos principios han probado ser insuficientes para la unificación, entonces Korsgaard tendría que renunciar a la tesis de la indispensabilidad de los principios; de lo contrario caería en la concepción dividida del sujeto. Aunque la autora proponga el modelo constitucional justamente como una manera de evitar el modelo de combate y la fragmentación de la agencia, la crítica que realizo de la mano de Corbí arroja un resultado contrario: el modelo consitucional reproduce y sigue comprometido con una concepción del yo dividido. Evidentemente, si el modelo de Korsgaard no es suficiente para unificar la agencia, la concepción de integridad (unidad) que ella defiende básicamente queda sin sustento.

6.2 ¿Primacía normativa de la moralidad?

El rigorismo moral kantiano, además de su compromiso con el modelo del yo dividido, está comprometido con lo que denomino la tesis de la *primacía normativa de la moralidad* (PNM). Esto significa que la moralidad o el deber, en el caso específico de la filosofía práctica de Kant, son normativamente superiores al resto de criterios normativos. Este punto ya lo he señalado repetidas

veces a lo largo de esta investigación. Incluso es una tesis que ya se encuentra presente desde el caso del asesino en la puerta: e el deber moral ha de prevalecer por encima de cualquier otra consideración práctica.

El punto fuerte del rigorismo clásico kantiano es que no hay excepciones posibles al deber perfecto. Korsgaard no se compromete explícitamente con una tesis tan fuerte. Sin embargo, al concebir la moralidad como una condición constitutiva de la agencia, la autora está señalando que renunciar a la moralidad significa renunciar también a la condición de agencia. Eso convierte a la moralidad en un compromiso inescapable, tan inescapable como lo es para Kant. En ambos escenarios, tanto el rigorismo kantiano, como el constitutivismo de Korsgaard están comprometidos con la tesis de la primacía normativa de la moralidad.

Lo que busco defender en §6.2 es que PNM es falsa. Mi crítica a Korsgaard se articulaen dos partes. En primer lugar, tendré en cuenta la versión fuerte de la tesis de Korsgaard: la idea de que la moralidad es incondicionada en virtud de que ella es constitutiva de la agencia. Denominaré a esta posición la *tesis de la primacía constitutiva de la moralidad* (§6.2.1). En segundo lugar, consideraré la versión débil de PNM: existe una jerarquía normativa y aunque la moralidad no sea incondicionada, sí es el criterio normativo superior o más importante en la deliberación práctica. Esta versión será denominada la *tesis de la prevalencia* (§6.2.2).

6.2.1 Prioridad constitutiva de la moralidad

La tesis de la primacía constitutiva de la moralidad es la consecuencia filosófica del constitutivismo de Korsgaard. El constitutivismo es una forma específica del constructivismo kantiano (Rawls, 1980) que ha sido adoptado recientemente en las discusiones sobre normatividad práctica (Bagnoli, 2017). La intuición principal del constitutivismo es que la normatividad no está fijada por hechos normativamente relevantes que son independientes de los agentes; al contrario, la normatividad práctica surge de las condiciones constitutivas de la agencia y de la acción. En otras palabras, el constitutivista encuentra que las condiciones que hacen posible la agencia y la acción son también la fuente de la normatividad:

> [P]ara saber lo que hace que un carro sea un buen carro, necesitamos entender qué son los carros, cuáles son sus funciones constitutivas […]. Análogamente, entonces, tal vez para saber cuáles acciones son buenas (o correctas, o soportadas por la razón, o racionales, o lo que sea), necesitamos una mejor comprensión de lo que son las acciones, o tal vez de lo que es ser un agente. (Enoch, 2006, p. 32)

La promesa del constitutivismo es que las condiciones constitutivas de la acción y de la agencia son también condiciones normativas, es decir, condiciones que sirven para determinar si una acción es correcta o buena. Korsgaard justamente se adscribe a este proyecto filosófico. En el capítulo anterior, intenté dar cuenta de estas condiciones constitutivas de la acción (imperativo categórico y ley moral) y de la forma en la que estas condiciones tienen valor normativo (razones públicas).

La ventaja del constitutivismo, al menos como yo interpreto el proyecto de Korsgaard, es que permite dar una respuesta al escepticismo normativo (Williams, 1981a) sin necesidad de comprometerse con ningún tipo de realismo. Contra el escepticismo, el constitutivismo establece verdades normativas objetivas; contra el realismo, estas verdades normativas no se justifican en hechos o cualidades objetivas independientes de los agentes. El constitutivismo propone fundamentar una normatividad universal y objetiva para la acción a partir de la misma constitución de la acción y la agencia humana[79].

¿En qué sentido el constitutivismo de Korsgaard implica la tesis de la primacía constitutiva de la moralidad kantiana? Como vimos en el capítulo anterior, el imperativo categórico (§5.2.3) y la ley moral (§5.2.4) son condiciones constitutivas de la acción y de la agencia, y además son condiciones que implican o representan compromisos morales; por tanto, en el marco del constitutivismo de Korsgaard, la moralidad se convierte en una condición de la agencia y de la acción. Cualquier criterio práctico no constitutivo está sometido a las condiciones morales constitutivas. Habría que agregar que, en la interpretación de Korsgaard, el imperativo categórico y la ley moral no son únicamente condiciones constitutivas de la acción y de la agencia. En realidad ellas son condiciones *auto*-constitutivas de la agencia. Esto se explica, recordemos, por la idea de que ser-agente no es

[79] Entre los filósofos que también han adoptado este proyecto se cuentan: Rawls (1980), Velleman (2004) y Rosati (2003), Bagnoli (2009).

una condición dada de antemano, tal y como lo concebía la tradición moderna. Ser-agente es un proceso, una tarea que se actualiza en cada acción, que a su vez está constituida por condiciones morales[80].

La posición de Korsgaard, como la interpreto, la obliga a comprometerse con la idea de que estas condiciones constitutivas no están, a su vez, constituidas por nada, pues ellas mismas auto-constituyen la acción. Que ellas sean condiciones auto-constitutivas las convierte en condiciones no condicionadas. En el mito hindú, por ejemplo, el mundo está sostenido por elefantes, y estos están sostenidos por una inmensa tortuga que flota por sí misma en el espacio. De la misma manera, la acción está constituida por condiciones morales que por sí mismas son suficientes y necesarias para configurar la agencia. No hay nada más básico que las soporte. Estas condiciones por sí solas auto-constituyen la agencia.

A partir de esta idea, resulta que la moralidad tiene una prioridad normativa primaria, pues es condición fundante y constitutiva de la acción y la agencia. En un sentido fuerte, esto significa que la moralidad es una condición incondicionada. No existe un criterio práctico más básico. ¿Qué ocurre, sin embargo si se cuestiona el carácter incondicionado de la moralidad?

a. Schmagencia [Schmagency] y el juego de la agencia

Una de las críticas más interesantes contra el constitutivismo de Korsgaard ha sido realizada por David Enoch (2006). La tesis de este autor es que es posible concebir otras formas posibles de agencia que no tengan las características constitutivas que Korsgaard les atribuye. *Schmagency* sería el nombre para designar este otro tipo de agencias.

> Clasifica mis movimientos corporales y a mí como quieras. Tal vez no pueda ser clasificado como agente sin tratar de constituirme a mí mismo. ¿Pero por qué debería ser un agente? Tal vez no pueda actuar sin aspirar a la auto-constitución, pero ¿por qué

[80] "El compromiso con la ley moral está incluido en la actividad en la que, por el hecho de ser humanos, estamos necesariamente comprometidos: la actividad de hacer algo de nosotros mismos. La ley moral es la ley de la auto-constitución, y como tal, es un principio constitutivo de la vida humana misma" (Korsgaard, 2009, p. xiii).

> debería hacerlo? Si tu razonamiento funciona [el de Korsgaard], esto demuestra que no me importan la agencia y la acción. Estoy perfectamente satisfecho con ser un "shmagente", un no-agente que es muy similar a los agentes pero que carece del objetivo (constitutivo de la agencia pero no de la "shmagencia") de la auto-constitución. Soy perfectamente feliz realizando "shmacciones" –eventos (no acción) que son muy similares a las acciones pero que carecen del objetivo (constitutivo de las acciones pero no de las shmacciones) de la auto- constitución. (Enoch, 2006, p. 41)

No me detendré a examinar a fondo la crítica de Enoch a Korsgaard. Sin embargo, me interesa analizar uno de los argumentos que, en mi opinión, captura el punto central de esa crítica. Enoch traza la siguiente analogía: en un juego de ajedrez existen unas reglas de juego. Si alguien no respeta estas reglas, literalmente no está jugando ajedrez, sino otro tipo de juego. Por tanto, las reglas del ajedrez son las que determinan si estoy o no jugando ajedrez. Se puede decir, entonces, que las reglas son constitutivas del juego de ajedrez. De la misma manera, la agencia, según Korsgaard, supone unas condiciones constitutivas, sin las cuales esta no sería posible. Si estas condiciones no se cumplen, literalmente se está renunciando a la agencia y a la acción: un individuo que renuncie a la moralidad, en sentido estricto, no está actuando, sus movimientos serían respuestas a estímulos o eventos que no podrían ser catalogados como acciones.

Enoch acepta que es posible que existan tales condiciones constitutivas para jugar ajedrez o para ser agente. En este sentido, el constitutivismo puede ser una estrategia correcta para fundamentar la normatividad. Sin embargo, el constitutivismo es limitado. Volviendo al caso del ajedrez, un jugador solo obedecería tales reglas (aceptaría la normatividad de las reglas) si *tiene una razón* para jugar ajedrez. Si, en cambio, el jugador no tiene tal razón, no tendrá razones para adoptar las reglas y tampoco tendrá razones para, por ejemplo, hacerle un jaque mate a su oponente. Más aun, sin una razón para jugar, el jugador bien podría estar frente al tablero y bien podría mover las fichas, y con todo, no se podría decir que está jugando ajedrez; estaría haciendo otra cosa.

Lo interesante de esta analogía es que las reglas constitutivas del ajedrez no nos dan, por sí solas, una razón para jugar ajedrez. La normatividad de las reglas puede ser constitutiva del juego, pero está limitada al juego mismo. Hace falta entonces una razón, ulterior y distinta a las reglas, para que el jugador juegue ajedrez. Si se traslada esta analogía al caso de la agencia, la situación es

semejante. Korsgaard postula unas condiciones constitutivas para la agencia. No cumplirlas implica renunciar a la condición de ser-agente. ¿Pero qué razón tenemos para jugar el juego de la agencia? Al igual que sucede en el caso del ajedrez, las condiciones constitutivas de la agencia están limitadas a una concepción específica de ser-agente (concepción determinada por el modelo constitucional). Pero estas condiciones, por sí mismas, no aportan una razón para ser-agente en el sentido que Korsgaard propone.

Una posible defensa de Korsgaard es decir que la analogía falla porque mientras jugar ajedrez es opcional, jugar el juego de la agencia es inevitable. No es posible elegir entre ser agente y no serlo. Para Korsgaard, la inevitabilidad de la agencia es un hecho fundamental. "Los seres humanos estamos *condenados* a elegir y a actuar" (Korsgaard, 2009, p. 1; énfasis en el original). El asentimiento reflexivo del que Korsgaard parte es una característica intrínseca de nuestra condición de seres humanos. De hecho, ella acepta, con algo de reserva[81], que su postura es naturalista, pues está fundamentada en hechos naturales como la autoconsciencia y el asentimiento reflexivo que se desprende de ella. Lo cierto, en todo caso, es que nuestra condición "natural" de autoconsciencia inevitablemente nos sitúa como agentes. De manera que si el juego de la agencia es inevitable, si no podemos retirarnos de él, ¿por qué habríamos de necesitar una razón para jugar el juego de la agencia?

Según Enoch, esta posible réplica no resuelve el problema que surge de la analogía del ajedrez.

> 'Tal vez', el escéptico de Korsgaard pueda decir, 'No puedo optar por no participar en el juego de la agencia, pero ciertamente puedo jugarlo a medias, incluso bajo protesta, sin aceptar como míos los objetivos supuestamente constitutivos'. El tipo de necesidad que tiene que tener el juego de la agencia para resolver el problema en la que nos encontramos es una necesidad normativa. Invocar otras necesidades aquí no resolverá el problema. Tal vez al darse cuenta de este problema, Korsgaard nunca tiene claro el tipo

[81] "He localizado finalmente la normatividad de la obligación en el hecho de la consciencia reflexiva […]. Mi concepción no depende de la existencia de seres sobrenaturales ni de hechos no naturales, y es consistente con la Concepción Científica del Mundo, si bien no forma parte de ella. En este sentido, [mi concepción] es una forma de naturalismo". (Korsgaard, 1996, p. 160).

de necesidad que invoca. (Enoch, 2006, p. 50; énfasis en el original)

Si entiendo bien a Enoch, se puede aceptar la inevitabilidad de la agencia, pero esto no significa que el agente esté comprometido o identificado con el juego de la agencia. Hace falta, además, una razón para jugar este juego, y esta razón, dice Enoch, ha de tener un carácter normativo para el agente. Lo que resalta el autor, pienso, es que la inevitabilidad del juego de la agencia no es una razón normativamente relevante para jugar la agencia. En otras palabras, la inevitabilidad de ser agente ("la condena a ser agente") no es una necesidad normativa, y por lo tanto, no se puede apelar a esta inevitabilidad para justificar el compromiso con el juego de la agencia.

Para fortalecer el punto de Enoch, quisera añadir que incluso en un escenario donde sea inevitable hacer algo, y el agente sea consciente de esta inevitabilidad, todavía es posible pensar que el agente tenga razones adicionales para hacer lo inevitable.[82] Esto significa que la inevitabilidad de un curso de acción o de un estado de cosas puede ser diferenciada de las razones que pueda tener el agente para hacer lo inevitable. El héroe que se enfrenta a una muerte inevitable puede morir porque lo concibe como su destino, o por el bien de la patria, o porque así defiende su causa. Difícilmente él muere porque 'es inevitable morir'. Asimismo, es inevitable jugar el juego de la agencia y esto implica que no podemos elegir jugar, pero no implica que la persona no tenga una razón adicional para jugar. De manera que si la razón para jugar el juego de la agencia no es intrínseca a las reglas constitutivas mismas, resulta que no es necesario estar comprometidos con las reglas constitutivas de la agencia, a pesar de que el juego de la agencia sea inevitable.

Es más, incluso si alguien llega a considerar que la inevitabilidad es una razón suficiente para jugar el juego de la agencia, resulta que esa persona todavía ha de adoptar esta inevitabilidad como su razón: 'juego porque es inevitable'. Esta idea marca una diferencia importante con respecto de lo que propone Korsgaard, pues adoptar la proposición 'juego porque es inevitable' como una razón

[82] Ángela Uribe cuestiona la naturaleza normativa de estas "razones adicionales". ¿A qué tipo de normatividad obedecen estas razones? Esta pregunta es fundamental porque pone de relieve la idea de que estas razones no tienen normatividad *moral*, y, sin embargo, como espero mostrar, sí obedecen a otro tipo de normatividad. Más adelante, cuando introduzca la distinción entre ética y moralidad, desarrollaré el concepto de normatividad ética, que es, en mi opinión, el tipo de normatividad al que podrían estar vinculadas estas "razones adicionales" (ver §6.2.2.).

implica entrar inmediatamente en un espacio normativo de razones que es de un orden distinto al *hecho* de que la agencia sea inevitable. La idea que me interesa resaltar, en todo caso, es que la razón para jugar el juego de la agencia es distinta (pertenece a otra categoría) a la normatividad de las condiciones constitutivas de la agencia. De manera que el constitutivismo no explica enteramente el fenómeno de la agencia, pues además de las reglas o condiciones constitutivas ha de existir una razón adicional para que el agente esté comprometido con su agencia. Y esta razón va más allá de las condiciones constitutivas propuestas por Korsgaard.

b. 'Razón para estar vivo' y moralidad

La conclusión que extraigo de la crítica de Enoch a la propuesta de Korsgaard es que el constitutivismo necesita apelar a un orden de razones externo a la propia agencia para justificar el juego de la agencia. Introduzco entonces una distinción entre condiciones constitutivas de la acción y condiciones extra-constitutivas. Korsgaard no admitiría esta distinción, puesto que las condiciones constitutivas son incondicionales con respecto a la agencia, y en esa medida son suficientes y auto-constitutivas de la agencia.

La tesis del constitutivismo, entonces, es correcta solo asumiendo que el agente tiene una razón extra-constitutiva para jugar el juego de la agencia, pues si no existe esta razón, ¿por qué habría de comprometerse el agente con unas reglas de un juego para el cual él no tiene una razón para jugar? En este sentido, hay algo que condiciona a las condiciones constitutivas de la agencia: que el individuo tenga razón para ser agente. Dicho esto, la idea que me interesa defender en lo que sigue es que una condición para que alguien juegue el juego de la agencia es que ese alguien *tenga una razón para estar vivo*. En otras palabras, lo que espero defender es que no es cierto que las condiciones constitutivas sean incondicionadas, pues estas están, a su vez, condicionadas a que se tenga una razón para vivir.

La idea que propongo, de hecho, se puede extraer de las propias palabras de Korsgaard. Y aunque ella no estaría de acuerdo con mi conclusión, creo que tendría que aceptarla. Como vimos, la autora introduce la noción de identidad(es) práctica(s). Estas identidades son fuente de obligaciones y deberes, y por lo tanto, son fuente de normatividad. Algunas de estas identidades prácticas son

incondicionales porque son con las que una persona más se identifica. Es por eso que traicionar alguna identidad incondicional supone la muerte práctica.

Korsgaard, sin embargo, también considera que hay otras situaciones que no están bajo nuestro control y que también pueden causar la muerte práctica. Una enfermedad, la invalidez o el dolor pueden "destruir nuestra identidad al destruir su fundamento físico, arrasar con nuestra memoria o imposibilitar el autocontrol" (1996, p. 201). Concuerdo con la autora en la idea de que en estas situaciones el suicidio "puede ser la única manera de conservar nuestra identidad y de proteger los valores por los cuales hemos vivido" (*ibid.*).

Existe un vínculo entre la pérdida de una identidad práctica y el suicidio. Cuando se ve aniquilada una identidad incondicional (como un proyecto fundamental, en las palabras de Williams), la persona puede tener una razón para el suicidio. Si hay algo que le importa a la persona, algo que se refleja en su identidad incondicional, y por alguna contingencia esta identidad se pierde o se elimina, entonces, digo, la persona puede considerar que el suicidio es la mejor forma de honrar aquello que ha perdido. Esta persona considera que no puede seguir hacia adelante si ya no puede vivir de acuerdo con la identidad incondicional que ha perdido.

Adicionalmente, continúa Korsgaard, existe otra forma de suicidio que se comete cuando la persona tiene una vida vacía de valor, cuando ya nada lo ata al mundo. Cuando esto ocurre, la persona tiene una razón para quitarse la vida. Este segundo tipo de suicidio es un fenómeno que demuestra la posibilidad de que exista una vida desposeída de cualquier valor, y esta clase de suicidio no es la negación de un valor en particular, "es la negación del valor mismo" (*cf.* Korsgaard, 1996, p. 202).

No estoy muy seguro de que la distinción entre estos dos tipos de valor sea muy precisa. En especial porque el primer tipo de suicida, el que ha perdido su identidad incondicional, llega a la misma situación del segundo tipo de suicida: ya nada tiene valor para ellos. Ambos comparten la misma razón para cometer el suicidio, o en otras palabras, el primer tipo de suicida fácilmente se convierte en el segundo tipo de suicida[83]. Para mi argumento, sin embargo, lo que interesa no es tanto la

[83] Agradezco esta observación a Ángela Uribe.

distinción, sino la relación entre el valor y el suicidio: una razón para quitarse la vida es la pérdida del valor o, en la versión más nihilista, la negación del valor.

Ahora bien, es llamativo que Korsgaard admita la posibilidad de la "negación del valor mismo". El argumento constitutivista de esta autora asume que un agente integrado admite el valor de la humanidad y, por tanto, habrá de admitir también el valor de la moralidad (5.2.3a-b). Y además, es esta identidad moral y esta identidad humana la que dota de valor al resto de identidades prácticas de una persona. Ella misma reconoce este punto: "[…] he sostenido que la obligación moral y el valor moral son condición de toda obligación y de todo valor" (Korsgaard, 1996, p. 202).

Sin embargo, la persona que cometa suicidio porque niega la idea misma de valor está negando a su vez el valor de la moralidad y el de la humanidad (Korsgaard, 1996, p. 202)[84]. Esta persona ilustra el caso de un escepticismo práctico absoluto. Y Korsgaard se pregunta: ¿hay algún argumento contra este escepticismo, una razón para no suicidarse cuando la vida misma está desposeida de todo valor? (1996, p. 203). Korsgaard admite que no existe un argumento contra el escéptico absoluto, porque "la vida misma es fuente de razones […] *el valor sólo existe si vale la pena vivir la vida*" (*ibid.;* énfasis añadido).

Esta respuesta confirma la idea que intento justificar. Extraigo de las palabras de Korsgaard la idea de que ni la humanidad ni la moralidad tendrían valor para alguien si ese alguien no tiene una razón para vivir. En otras palabras, tener una vida a la que se le atribuye sentido es la fuente de todo valor, incluyendo el valor de la humanidad y el valor de la moralidad. Traduciendo esta idea a la discusión sobre constitutivismo, resulta que las características constitutivas de la agencia (identidad humana, identidad moral, imperativo categórico, ley moral) están condicionadas a que el agente tenga una razón para jugar el juego de la agencia; y una razón necesaria para ser agente, insisto, es que la persona tenga una razón para estar vivo.

Korsgaard, por supuesto, no podría aceptar esta conclusión que va en contra de su constitutivismo.

---

[84] Williams va en la misma dirección: "No veo por qué tendría que considerarse como una derrota de la razón o de la racionalidad la impotencia frente al estado de este hombre [que busca suicidarse]; su estado representa más bien una derrota para la humanidad" (Williams, 1972, p. 4).

Su respuesta, de hecho, consiste en despojar de humanidad al escéptico radical. Para ella, el escepticismo radical, de ser concebible, supondría vivir una vida no humana. ¿Acaso –se pregunta Korsgaard– el escéptico radical debe suicidarse? Y su respuesta es que no debe hacerlo porque hay un sentido en que él no está obligado a nada, su vida está desposeída de todo valor y sin valor no hay obligación. En el mejor de los casos, esa vida sería un aglomerado de deseos ante los que este solo respondería momentáneamente. No existe ni puede existir ningún principio que pueda guiar sus actos. El escéptico radical, entonces, lleva una vida muy distinta a la del resto de seres humanos. El precio de esta forma de vivir, dice Korsgaard, es la renuncia a la condición de humanidad (1996, p. 202), un resultado apenas obvio, si es verdad que esta forma de escepticismo renuncia al valor de la identidad humana y al valor de la moralidad.

Para los intereses que persigo, sin embargo, no importa tanto si el escéptico radical pierde su humanidad, o si su vida puede ser concebida como una vida humana. No me interesa salvar al escéptico radical. Lo que me interesa señalar es que Korsgaard está asumiendo que el valor de la humanidad y el de la moralidad dependen o están condicionados a que la persona tenga una razón para vivir. La humanidad y la moralidad no son valores que necesariamente tenga una persona. Las personas no están necesariamente sujetas u obligadas a valorar su humanidad o la moralidad. Y lo que es más interesante: si alguien no tiene una razón para vivir, tampoco tendrá una razón para cumplir la moralidad.

Pensemos en el pasaje de la *Fundamentación de la metafísica de las costumbres* en el que Kant condena explícitamente el suicidio. Según él, el suicidio es una violación de un deber perfecto contra uno mismo (*F* 4:397) (ver el análisis en §2.3). Aun cuando la persona se encuentre en una situación de padecimiento o esté viviendo un gran infortunio, y por eso haya perdido todo gusto por la vida, sigue siendo un deber moral conservar la vida[85]. El escenario que ilustra Kant es el de una persona que padece una desgracia porque ha perdido sus bienes. Esto supone que la persona valora estos bienes, y que su pérdida representa un gran infortunio. Estos bienes o valores son tan importantes que la persona está dispuesta a abandonar la vida porque considera que sin ellos no vale la pena

[85] "[…] [C]uando los infortunios y una pesadumbre desesperanzada han hecho desaparecer por entero el gusto hacia la vida, si el desdichado desea la muerte, más indignado con su destino que pusilánime o abatido, pero conserva su vida sin amarla, no por inclinación o miedo, sino por deber, entonces alberga su máxima un contenido moral" (*F* 4:398).

seguir viviendo. Como interpreto esta situación, el suicida piensa que *si* pierde X, entonces nada más puede ser valorado por él y, por tanto, ya no vale la pena seguir viviendo.

El suicida de Kant concibe el suicidio porque ha perdido aquello que valora, pero qué pasaría si, por ejemplo, como en el caso de Job, recuperara los bienes perdidos. Es probable, creo, que esta persona ya no tendría motivos para cometer suicidio porque su consideración sobre cometer suicidio está condicionada a la pérdida de aquello que ha perdido. Si recupera X es probable que recupere el valor de aquello que lo mantiene atado a la vida. Se trata de la misma idea de Schopenhauer: "*el suicida quiere la vida*, simplemente está insatisfecho con las condiciones en que se le presenta" (*WWV* 471; énfasis añadido). Esta idea aparentemente paradójica tiene sentido si se piensa que la persona que contempla la posibilidad de suicidio lo hace porque esta vida, su vida, no vale la pena ser vivida, pues está padeciendo o sufriendo una vida que no se acerca a lo que, para ella, es digno de ser vivido. La persona valora tanto la idea de una vida valiosa que termina por renunciar a su mala vida. De la misma manera, el suicida kantiano puede recuperar "el gusto por la vida" si tuviera la fortuna de recuperar aquello que valora.

Normalmente, sin embargo, quien ha padecido grandes infortunios durante un largo tiempo pierde toda esperanza y todo "gusto por la vida" y no encuentra ya más sentido en la vida en general. Tal es la condición del escéptico radical del que hablé antes, alguien para quien la vida carece de todo valor. De hecho, como señalé antes, el suicida que perfila Kant puede llegar a convertirse en un escéptico radical. Su desgracia es tan profunda y prolongada que su vida ya no tiene valor, y hasta llegue a pensar que cualquier vida humana –en general– carece de valor porque asume que la condición humana es intrínsecamente un sinsentido y, por tanto, no vale la pena vivirla.

Llegados a este punto, no es obvio que el escéptico radical tenga el deber moral de conservar la vida. La posición de Kant funciona asumiendo que todavía existe algo que pueda llegar a atar al agente al mundo, todavía existe algo valioso para esta persona (p.e. la esperanza de recuperar lo perdido, o la esperanza de una vida mejor), y en esa medida todavía está vinculado a cumplir la moralidad porque, como dice Korsgaard, el valor moral es condición de todo valor. El escéptico radical, en cambio, cuestiona la fuerza de este argumento: si nada le importa, él no está obligado a nada, ni siquiera a conservar su vida. Quien no puede atribuir un valor a nada, quien no tiene

ninguna razón para vivir, tampoco está atado a la moralidad.

El mismo Kant no está muy alejando de la anterior idea. En la *Metafísica de las costumbres*, él sostiene que "aniquilar el sujeto de la moralidad en la propia persona equivale a erradicar del mundo la existencia de la moralidad misma" (*MC* 422-423). Con estas palabras, él está mostrando la incompatibilidad entre la moralidad y la idea del suicidio. Y desde su perspectiva, esta incompatibilidad muestra la inmoralidad del suicidio. Pero en realidad, esta conclusión solo es válida asumiendo que la moralidad tiene primacía normativa incuestionable. Sin embargo, cuando el escéptico radical entra en este escenario, gustosamente aceptaría las palabras de Kant, solo que para él estas palabras no expresan la inmoralidad del suicidio, sino algo más prufundo: la idea de que el suicidio es un rechazo a la idea misma de valor moral.

Pensemos mejor esta idea con la siguiente historia. En 1791, Kant recibió la carta de María von Herbert, una mujer austriaca muy entusiasta de la filosofía kantiana. En la carta, ella le expresaba a Kant la siguiente preocupación:

> De la Señorita María von Herbert a Kant
>
> Agosto de 1971
>
> Gran Kant:
>
> Como un creyente pide a Dios, yo le pido ayuda, consuelo para prepararme para la muerte. Sus textos me han probado que existe una vida futura. Pues para esta vida, no he encontrado nada, absolutamente nada que pudiera reemplazar el bien que he perdido, ya que he amado a alguien quien, a mis ojos, albergaba dentro de él todo lo que es valioso, así que yo viví solo por él, todo lo demás en comparación eran solo desperdicios, baratijas. Pues bien, yo he ofendido a esta persona, debido a una mentira por mucho tiempo guardada, que ahora le ha sido descubierta, aunque en ella no hay nada desfavorable en contra de mi carácter, yo no tengo ningún vicio en mi vida que necesite esconderse. La mentira le bastó para que su amor se desvaneciera. Como un hombre honorable, no rechazó mi amistad. Pero ese sentimiento interno que una vez, espontáneamente, nos dimos el uno al otro, ya no existe —oh, mi corazón se astilló en

> mil pedazos. Si no hubiera leído su trabajo ciertamente habría puesto fin a mi vida. Pero la conclusión que extraigo de su teoría me detiene —es incorrecto morir porque mi vida sea un tormento, más bien se supone que debo vivir en virtud de mi ser. Ahora póngase en mi lugar, y condéneme o consuéleme. He leído la *Metafísica de las costumbres*, y el imperativo categórico, y no me ayudan ni un poco. Mi razón me abandona justo cuando más la necesito. Respóndame, se lo imploro — ¿o no actuará de acuerdo con su propio imperativo? […]. (C 11:273)

No reproduciré la extensa respuesta de Kant a esta carta (primavera 1792) (*C* 11:331-334). En líneas generales, el autor le expresa a von Herbert la intención que tiene de ayudarla a través de un "sedativo moral puro". Le dice que la amistad o el matrimonio suponen un respeto recíproco. La amistad ideal supone una "apertura de corazones sin reserva". Sin embargo, la debilidad humana tiende a limitar esta apertura y a guardarse "parte de los propios pensamientos dentro de uno". No podemos esperar la franqueza del otro, dice Kant, porque "las personas tienen miedo de revelarse enteramente ante los demás". No obstante, la falta de franqueza no es lo mismo que la deshonestidad, pues quien no es franco no miente, solo no dice toda la verdad. La deshonestidad, en cambio, es una mentira y supone una violación al deber para con uno mismo. Más adelante, en su respusta, Kant le pregunta a von Herbert si su malestar surge de "la imprudencia de confesar o de la inmoralidad intrínseca de la mentira" (*C* 11:333). En el primer caso, no hay un motivo moral en juego, porque el dolor viene de las consecuencias de la confesión. En el segundo caso, dice Kant, no hay razón para alejar este malestar moral de la propia mente (se trata de una suerte de malestar legítimo por violar el deber moral). Finalmente, Kant dice que si el amado de von Herbert se aleja es porque su amor no era verdadero, era un amor físico y no moral; habrá entonces que darle tiempo para que convierta su frialdad en un "amor firme". Si se trata de un amor falso, entonces habrá que aceptarlo y enfrentarlo con "compostura". El filósofo termina su carta con las siguientes palabras: "Aquí, entonces, querida amiga, encontrará las acostumbradas divisiones de un sermón: instrucción, castigo y consuelo. Entréguese a las dos primeras, cuando ellas tengan su efecto, el consuelo será encontrado por sí mismo".

En una segunda carta (enero de 1793), von Herbert nuevamente pide ayuda a Kant, solo que esta vez, le expresa un profundo malestar. Ella empieza agradeciendo al filósofo por su respuesta y le señala que se encuentra en un dilema: la amistad requiere de la honestidad, pero si la verdad fuera

descubierta, la amistad se terminaría; de manera que, a sus ojos, la verdad y la amistad estaban en tensión. Von Herbert escribe estas elocuentes palabras:

> Mi visión es clara ahora. Siento que un gran vacío se extiende dentro de mí, y sobre todo alrededor mío —así que casi me encuentro a mí misma innecesaria, superflua. Nada me atrae. Estoy atormentada por un aburrimiento que hace mi vida intolerable. No piense que soy arrogante por decir esto, pero las demandas de la moralidad son muy fáciles para mí. Con entusiasmo haría dos veces lo que ellas ordenan. Ellas solo obtienen su prestigio del atractivo del pecado, y no me cuesta esfuerzo resistirlo. (C 11:401)

Von Herbert continúa diciendo que es un consuelo saber que la moralidad está ligada a la sensualidad, porque cuando muera, no tendrá que estar atada a esta "vida vegetativa y vacía" ni a estas "ligeras tareas de la moralidad". Ella aprende de la experiencia que a la mayoría de las personas les "gusta vivir". Y es probable, como ella misma lo reconoce, que la causa de su "desviación" sea su débil salud. Lo cierto, sigue von Herbert, es que solo quiere una cosa: "acortar esta vida sin sentido, una vida que estoy convencida no será ni mejor ni peor". Finalmente, le pide a Kant que examine esta cuestión, "porque mi concepción de la moral guarda silencio aquí, mientras que habla decididamente en todo lo demás […]" (*C* 11:403).

Von Herbert, claramente, ha perdido interés en la vida. Ella quiere morir y no encuentra apoyo en la moralidad kantiana. Lamentablemente Kant no responde a esta segunda carta. Las razones las desconozco, pero, a la luz de la discusión que aquí planteo, puedo suponer que Kant rechazaría la posibilidad del suicidio en el caso de von Herbert, como en cualquier caso, porque lo considera contrario al deber perfecto contra uno mismo, un deber perfecto.

Por lo que he venido diciendo, resulta, sin embargo, que el deber de conservar la propia vida solo tiene sentido si se asume que se tiene alguna razón para ser moral. ¿Qué sucede, sin embargo, con el caso de Maria von Herbert? Nada tiene valor para ella. Su amor no correspondido la ha llevado a una depresión. Su carta expresa en varios momentos un profundo malestar y una pérdida en interés por la vida. Hay algo aún más inquietante en las palabras de von Herbert: la moralidad ya no le sirve como una razón para seguir con su vida. El argumento de que el suicidio atenta contra el deber no es algo que le ayude a encontrar una razón para seguir viviendo.

Este es nuevamente el punto ciego del constitutivismo que he tratado de delimitar atrás: aunque, según Korsgaard, el juego de la agencia requiera compromisos con la moralidad, he intentado mostrar que hace falta, además, una razón para jugar este juego. Usualmente esta razón parece darse por descontada, prueba de ello es que las personas no suelen experimentar una pérdida de sentido en sus vidas, y las contingencias de la vida y las múltiples actividades impiden incluso plantear la pregunta por las razones para continuar viviendo. Sin embargo, cuando se vive una crisis de sentido como la que vivió von Herbert, aparece el límite del compromiso con la moralidad.

c. ¿Primacía constitutiva de la moralidad?

La carta de Maria von Herbert ha expresado una crítica a la filosofía práctica kantiana muy profunda, pero lastimosamente Kant no la supo entender. Susan Wolf, en su excelente artículo "Meaning and Morality" (1997), ha logrado sintetizar el corazón de esta crítica de una manera muy clara:

> La situación de la persona deprimida revela algo sobre la situación de las personas normales, quienes no tienen dificultad en encontrar razones para vivir o para tener interés en el mundo. Reflexionar en quienes padecen una depresión profunda nos hace reconocer que las preguntas '¿por qué es mi responsabilidad mejorar el mundo, o evitar que se deteriore?', '¿por qué debería intentar que el mundo sea un mejor lugar?' pueden ser preguntas legítimas, y no, como tiende a imaginar el moralista, simples protestas inauténticas de una naturaleza egoísta y avara. (Wolf, 1997, p. 308)

Una persona que no atribuye ningún sentido en su vida, que está desposeída de valor y de interés en el mundo, tampoco tiene ninguna razón para comprometerse con la moralidad. "Al abandonar lo que le da una razón para vivir y para preocuparse por el mundo, la persona renuncia también a tener una razón para preocuparse por el orden moral" (Wolf, 1997, p. 313). El punto relevante de la crítica es que no es evidente que la moralidad sea siempre y necesariamente el valor fundamental e incondicionado del cual se desprenden los demás valres o compromisos.

En la lectura rigorista de la filosofía kantiana que he defendido en esta investigación, el agente debe

cumplir siempre el deber moral. Korsgaard, si bien admite la posibilidad de los conflictos prácticos y admite también que la moralidad no es el único valor en la vida humana, todavía está comprometida con la idea de que la moralidad tiene prioridad normativa sobre cualquier valor o criterio práctico no moral. La pregunta que he estado tratando de articular hasta acá es: ¿por qué una persona debería aceptar la prioridad normativa de la moralidad? Para el caso de la persona que no encuentra sentido a su vida y que no tiene interés en el mundo, parece que no hay una respuesta satisfactoria a esta pregunta (*cf.* Wolf, 1997, p. 308). Es esto lo que pretendo ilustrar con el caso de Maria von Herbert.

Que la moralidad no ofrezca una respuesta para la crisis de sentido, que tampoco ofrezca una razón para vivir, implica un serio cuestionamiento a la tesis de la primacía constitutiva de la moralidad. El punto de Wolf, que comparto, es que la moralidad no es un valor incondicionado. Por tanto, no es verdad que la moralidad tenga una primacía constitutiva garantizada.

> [E]s irrazonable que la moralidad necesariamente gane en caso de un conflicto con la comprensión última que tenga la persona sobre el sentido, sin pensar que esto cuestiona el contenido de la moralidad. Uno podría en cambio revisar la propia perspectiva acerca de la fuerza racional de la moralidad. (Wolf, 1997, p. 310)

d. Reivindicación de la crítica de Williams: moralidad y sentido

Puede ser cierto entonces que la moralidad sea una condición constitutiva la agencia, como defiende Korsgaard, pero la moralidad, a su vez, está condicionada a que la persona tenga una razón para vivir. Por lo tanto, la incondicionalidad de la moralidad que pretende el constitutivismo es falsa. Esta conclusión revive la crítica de Williams a la moralidad kantiana. En el capítulo 5, quise mostrar la respuesta que ofrece Korsgaard al problema de la integridad: si la moralidad es una condición auto-constitutiva de la agencia, entonces esta tiene prioridad normativa por encima de los proyectos fundamentales. Lo que aquí he tratado de hacer es cuestionar la primacía constitutiva de la moralidad, o en otras palabras, he tratado de cuestionar la idea de que la moralidad es incondicional en virtud de su condición constituyente de la agencia.

Según mi exposición en el capítulo 4, coincido con Wolf en que la crítica de Williams no debe entenderse en términos de una oposición entre moralidad y felicidad, sino en términos de una oposición entre moralidad y el sentido (*meaning*) que una persona tiene de su vida (*cf.* Wolf, 1997)[86]. De hecho, una amenaza a la integridad no significa que la moralidad atente contra aquello que nos hace felices, sino contra nuestros proyectos fundamentales, es decir, contra aquello que dota de sentido nuestras vidas y nos da una razón para estar vivos.

Al insistir en la diferencia entre felicidad y sentido, y al leer la crítica de Williams a través de la categoría de sentido, el problema de la integridad que plantea Williams gana mucho más peso: ahora no solo se trata de una crítica a las demandas impersonales de la moralidad kantiana, también se trata de un rechazo a la tesis de la incondicionalidad de la moralidad:

> Williams está en lo cierto al señalar que las formas de moralidad kantiana no aciertan lo suficiente en reconocer el hecho de que nuestro compromiso con la moralidad es racionalmente […] condicionado –condicionado […] no a que el agente mantenga un grado de felicidad o tenga la oportunidad para perseguirla, sino [condicionado] a que mantenga una conexión con algo que haga su vida significativa, que le dé [a las personas] una razón para vivir y una razón para preocuparse por lo que pasa en el mundo. (Wolf, 1997, p. 310)

Insisto en que el problema de la integridad que plantea Williams no enfrenta la moral a la inclinación, sino a los proyectos fundamentales, a aquello que nos da una razón para seguir adelante. Si esto es así, ¿cómo es posible sostener la incondicionalidad de la moralidad, aun cuando ella implique renunciar a las razones que tenemos para vivir? Con la figura del escéptico radical y con la hostoria de Maria von Herbert, he tratado de mostrar que la incondicionalidad de la moralidad se viene abajo cuando alguien no tiene una razón para jugar el juego de la agencia porque sencillamente no tiene una razón para estar vivo. ¿Qué repercusiones tiene cuestionar la

[86] Wolf insiste en la importancia de distinguir la felicidad y el sentido (1997, p. 300ss). No se trata de lo mismo. Pueden existir vidas felices (una vida sin necesidades y con bienestar) sin sentido, y también puede haber vidas infelices con mucho sentido, pues es verdad que las personas están dispuestas a sacrificar su bienestar por perseguir aquello que les da una razón para estar vivas (nuevamente, pensemos en los héroes o los mártires).

incondicionalidad de la moralidad para la posición constitutivist de Korsgaard? Básicamente, la idea es que al negar la incondicionalidad de la moralidad se debilita la estrategia constitutivista de Korsgaard. Según esta autora, si la moralidad es una condición auto-constitutiva de la agencia, naturalmente esta tendrá prioridad normativa. Sin embargo, como he mostrado, el compromiso y el valor de la moralidad dependen de que la persona tenga una razón para vivir. De manera que incluso si el constitutivismo de Korsgaard es correcto, no puede sostener la incondicionalidad de la moralidad.

Hay otra forma de ver esta misma idea. Según mi interpretación, la tesis de la incondicionalidad de la moralidad kantiana viene acompañada de una "perspectiva impersonal" de la racionalidad práctica. Esto es evidente tanto en Kant (yo nouménico), como en Korsgaard (identidad moral). Dicho de otra manera, la incondicionalidad de la moralidad delata el carácter impersonal de las exigencias morales. Pues bien, concuerdo con Wolf cuando señala que Williams está atacando justamente la idea de que la "perspectiva impersonal" es fuente de razones absolutas e incondicionales: no puede haber garantía de que las razones impersonales siempre tendrán la autoridad para vencer a las razones que nacen de nuestro interés de estar en el mundo (*cf.* Wolf, 1997, p. 314). Por el contrario, si la tesis de la primacía constitutiva de la moralidad es falsa, es decir, si "nuestra susceptibilidad a razones generadas por el punto de vista impersonal está condicionada a nuestra capacidad para tener intereses y razones que no son impersonales" (*ibid.*), entonces hay lugar para que razones personales no necesariamente se depongan ante razones morales e impersonales.

6.2.2. Tesis de la prevalencia

Si la moralidad no es un valor incondicional en el sentido que supone el constitutivismo de Korsgaard, entonces la tesis de la primacía constitutiva de la moralidad es falsa. Esta tesis, como ya había señalado, representa una defensa fuerte para justificar la primacía normativa de la moralidad. No obstante, aun si la tesis constitutiva de la moralidad es falsa, eso no significa que se deba renunciar a la idea de que la moralidad tenga primacía normativa. De hecho, es pensable todavía una versión moderada para defender la prioridad normativa de la moralidad. Un defensor de esta versión sostendría que aunque la moralidad no es incondicional, sí es el criterio práctico más importante, y por tanto, ha de imponerse por encima de los criterios no-morales. Tomando

prestada la expresión de Corinna Mieth, esta es la tesis de la prevalencia [*overridingness-thesis*]: "si la moralidad está en conflicto con otros valores […] la moralidad debe siempre prevalecer –esta es la esencia de su fuerza normativa incondicional" (Mieth, 2017, p. 242).[87]Aunque no conozco una defensa explícita de esta tesis, sí considero que no son pocos los autores(as) que se adhieren a ella.[88] De hecho, la tesis de la prevalencia es la posición más natural para aquellos que buscan mantener la primacía normativa de la moralidad sin comprometerse con el rigorismo kantiano.

La tesis de la prevalencia, como la interpreto, no supone un compromiso rigorista con la moralidad y tampoco elimina la posibilidad de que existan otras fuentes de valor distintas a la moralidad. Para un defensor de esta tesis bastaría con mostrar que, frente a un conflicto práctico, la moralidad tiene prioridad normativa y debe imponerse sobre el resto de criterios. En todo caso, esta tesis es suficiente para sostener una forma moderada de la prioridad normativa de la moralidad.

La tesis de la prevalencia tiene la ventaja conceptual de aceptar los límites de la moralidad sin renunciar a su prioridad normativa. El defensor de esta tesis afirma que la moralidad es el criterio práctico más importante, excepto en los casos excepcionales en los que las razones para vivir estén amenazadas (una forma de amenaza a la integridad). Otra forma de entender este punto sería diciendo que el defensor de la tesis de la prevalencia puede aceptar el problema de la integridad y, por lo tanto, reconoce que la moralidad no es un valor absoluto e incondicional. Y con todo, diría este defensor, la moralidad tiene prevalencia siempre que no amenace nuestros proyectos fundamentales o aquello que nos da una razón para vivir.

Desde esta perspectiva, entonces, es posible aceptar la crítica a la incondicionalidad de la moralidad y, sin embargo, replicar que esto no afecta la prioridad normativa de la moralidad, pues esta solo se vería superada cuando entre en tensión con el proyecto fundamental que alguien tiene;

---

[87] Mieth introduce la expresión "tesis de la prevalencia" para referirse a la prioridad normativa que tiene la moralidad cuando se enfrenta a la felicidad [*happiness*]. Por mi parte, en esta investigación no busco sopesar el conflicto moralidad-felicidad, sino aquel que se da entre moralidad y sentido [*meaning*]. Tomo prestada su expresión pues me parece muy descriptiva y precisa, pero la uso en un sentido distinto.

[88] Autoras como Barbara Herman (1993a) e incluso la misma Korsgaard en "The Right to Lie" (2000b) (mucho antes antes de *Self-constitution* (2009)), pueden interpretarse a la luz de esta tesis. Asimismo, autores como Paton (1947), Hill (1992), Schapiro (2006) y Cholbi (2009), por ejemplo, intentan compatibilizar el imperativo categórico con las excepciones al deber, y en esta medida aceptan que la moralidad tiene prevalencia, aunque no en todos los casos.

pero en el resto de casos la moralidad tiene prioridad normativa. Además, si es verdad, como señala Susan Wolf, que los casos en los que existe una amenaza al proyecto fundamental que alguien tiene son raros (*cf.* 1997, p. 301), entonces en la cotidianidad de la vida práctica de los seres humanos la moralidad tendrá prioridad normativa.

En esta última sección me propongo justamente cuestionar la tesis de la prevalencia. Me interesa defender que existen otras fuentes de valor que no necesariamente son morales y que pueden llegar tener prioridad normativa por encima del valor moral. De esta manera, no estoy sugiriendo que la moralidad no pueda tener prioridad normativa, más bien, sugiero que la moralidad no siempre tiene prioridad normativa. Esta sección está divida en cuatro partes. En la primera, introduzco la distinción de Williams entre ética y moralidad que me servirá como marco de interpretación para realizar la crítica a la tesis de la prevalencia. En la segunda parte, me concentro en los conceptos de importancia y prioridad deliberativa; ambos pertenecen al ámbito de la ética y son ejemplos paradigmáticos de criterios normativos no morales. La tercera parte la dedico a la noción de necesidad práctica. Con esta noción intento caracterizar un fenómeno práctico que sirve de contrapeso normativo a la moralidad. Finalmente, en la última parte, argumento que la moralidad kantiana tiene que aceptar criterios normativos no morales y, por ello, debe renunciar a la tesis de la primacía normativa de la moralidad. Esta conclusión, naturalmente, representa un duro golpe para la imagen kantiana de la moralidad, ¿qué quedaría de esta imagen si se rechaza la idea de que la moralidad es el valor práctico más importante?

a. La distinción entre ética y moralidad

Sócrates creía que la filosofía tenía la capacidad de responder la pregunta por cómo dirigir la propia vida. Él esperaba que la filosofía nos diera una "comprensión general y abstracta, racionalmente reflexiva" de la vida humana (Williams, 2006c, p. 1). La pregunta socrática tiene la característica de atender a las preocupaciones más serias y propias de toda vida humana y da en el blanco respecto de algo que concierne a todos los seres humanos. La pregunta, tal como fue formulada por Sócrates, tiene la virtud de haber sido planteada en un estado de infancia teórica. Ella es un punto de partida. De ahí que sea una forma de "empezar de nuevo, interrogando reflexivamente al sentido común o a nuestras preocupaciones morales o éticas sin el peso de los textos ni de la tradición de los estudios filosóficos" (Williams, 2006, p. 2). Hoy ya no estamos en tal estado de

inocencia teórica, y el peso de la tradición de una u otra manera se ha impuesto hasta el punto de que no es claro que el interés práctico y el espíritu que tenía la pregunta socrática se mantengan actualmente. De hecho, la pregunta de Sócrates ha perdido su fuerza, y la historia de las ideas ha visto surgir otra pregunta de naturaleza muy distinta: qué se debe hacer. A esta la llamaré la pregunta kantiana (*C1* A805 B833).

La distinción entre ética y moralidad puede trazarse a partir de la diferencia entre la pregunta socrática y pregunta kantiana. El primer rasgo de esta diferencia radica en que la pregunta de Sócrates es más primaria que la pregunta kantiana. Me refiero a que la pregunta de Sócrates está abierta a distintas y variadas respuestas sobre cómo vivir. En esta medida, esta pregunta es más receptiva y amplia con respecto al espectro de sus posibles respuestas. Por ejemplo, es posible decir que se ha de vivir de acuerdo con el placer, otros dirán que se ha de vivir buscando la virtud, otros el deber, o a la *ataraxia*. El reto estaría en mostrar cuál, entre varias posibilidades, es la respuesta correcta. En constraste, la pregunta kantiana es más restrictiva respecto al espectro de posibles respuestas; y esto se debe a que esta ya asume, de entrada, que la vida práctica de los seres humanos está guiada por la obligación, es decir, por lo que debemos hacer. ¿En qué sentido entonces la pregunta socrática es más primaria que la kantiana? De nuevo, la pregunta socrática abre la posibilidad a respuestas diversas, y una respuesta posible a la pregunta sobre cómo se ha de vivir es recurriendo a la deontología kantiana: hemos de vivir siguiendo el deber. Lo que busco expresar es que la pregunta kantiana se plantea después de asumir una respuesta respecto a la pregunta socrática: dado que se ha de vivir con el propósito de cumplir con el deber, dirá el kantiano, la pregunta que se impone es cómo debemos actuar. Esto significa que la pregunta deontológica de Kant ("cómo debo actuar") solo es posible si previamente se ha aceptado que la vida humana ha de estar guiada por el deber.

Ambas preguntas tienen un interés práctico, y las respuestas que demos a ellas serán formas de regular la vida y la acción humanas. En este sentido, las dos preguntas son preocupaciones prácticas —no teóricas–, y sin embargo son de naturaleza muy diferente. Tal y como me interesa plantearlo, diría que la diferencia radica en que la pregunta de Sócrates es de naturaleza ética, mientras que la pregunta kantiana es de naturaleza moral. ¿En qué consiste entonces la distinción entre ética y moralidad? Esta distinción podría trazarse de varias maneras, pero para los propósitos de esta

investigación, sigo la propuesta de Bernard Williams (2006c).

Williams no tiene una definición precisa del término *moralidad*, él, más bien, ofrece caracterizaciones repartidas en uno u otro lugar de su obra. Tal vez sea en el artículo "Morality, the Peculiar Institution" (2006b) donde el autor intenta caracterizar de mejor manera el concepto de moralidad. Como es sabido, el autor es muy crítico y escéptico respecta a la moralidad; de hecho, el objetivo de su artículo es explicar "en qué sentido estaríamos mejor sin la moralidad" (Williams, 2006b, p. 174).

Independientemente de la posición de Williams respecto a la moralidad, lo importante por ahora es entender que, a pesar de la escueta caracterización que él hace del término, es evidente que traza una jerarquía entre moral y ética: la moralidad es un tipo específico de pensamiento ético, la moralidad es "un sistema especial, una variedad particular de pensamiento ético" (Williams, 2006b, p. 174). A la moralidad la caracteriza"su espíritu, sus objetivos subyacentes y la representación general de vida ética que esta implica" (*ibid.*). La particularidad de la moralidad radica en el uso del concepto de obligación para delimitar su representación de la vida ética. No se trata de una noción general y hasta cotidiana de obligación[89]; en realidad, la moralidad supone un concepto de obligación especial, se trata de la obligación moral.

Para Williams, una de las características de la moralidad es tratar de convertir y reducir todo a obligaciones morales (Williams, 2006b, p. 180). Desde la perspectiva de la moralidad kantiana, la noción de obligación moral es considerada fundamental y básica, y por ello el resto de conceptos prácticos tendrían que poder reducirse o depender de ella. ¿Cómo se lleva a cabo este reduccionismo? Para Williams, la respuesta a esta pregunta es que una obligación particular se extrae de una obligación más general. Este es el "principio de una obligación, otra obligación"[90] [*the obligation-out, obligation-in principle*] (Williams, 2006b, p. 181). El principio da lugar a lo que denomino un *reduccionismo deontológico,* pues toda acción particular es traducida a una obligación particular en virtud de que está respaldada por una obligación general. En el caso particular del

[89] Más adelante, cuando introduzca la noción de necesidad práctica (§6.2.2c), volveré sobre una noción más amplia y general de la obligación práctica.
[90] Tomo prestada esta expresión de la traducción al español que hace Sergi Rossell de *Ética y los límites de la filosofía* (2016, p. 223).

sistema moral kantiano, el principio del respaldo de la obligatoriedad es el imperativo categórico: cualquier máxima de acción que sea moral está siempre respaldada por este imperativo. Asimismo, una máxima que no cumpla con el requisito de universalización del imperativo moral es una máxima no moral. El universo de la moralidad está siempre restringido al imperativo categórico, nada que no tenga su respaldo puede tener la impronta moral. A través del reduccionismo deontológico, la filosofía kantiana puede justificar su dominio y monopolio sobre la vida moral humana. De esta manera, todos los ámbitos de la moralidad estarían dentro del universo de la obligación, la economía práctica usaría una moneda común: la obligación moral[91].

Con esta caracterización se destaca otro rasgo de la distinción entre ética y moralidad. Por moralidad, siguiendo a Williams, entiendo una perspectiva ética que está determinada por la obligación moral. Según esta perspectiva, la descripción de la acción correcta ha de darse exclusivamente en términos de obligación o deber moral. No existe una acción correcta al margen o independientemente del deber moral; análogamente, una acción incorrecta es aquella que representa una violación de un deber moral.

La ética, en contraste, y como ya he indicado, es una categoría más amplia que la de moralidad, una categoría que ha de ser concebida a la luz de la pregunta socrática y, por lo tanto, atiende a una preocupación por cómo vivir la vida antes que una preocupación por cómo se debe actuar. Al ser la ética una categoría tan amplia, es difícil dar cuenta con precisión de todos los elementos y fenómenos éticos que ella comporta, y resulta más difícil diferenciarlos de los no-éticos (Rivera, 2007, p. 76). De hecho, interpreto esta vaguedad como un rasgo importante del universo ético, pues es una muestra de su pluralidad y riqueza. Tal vez esto explica por qué Williams tampoco se ocupa de definir la ética, aunque sí caracteriza algunos de sus rasgos sobre los que vale la pena ampliar.

---

[91] Esto no significa que en la tradición kantiana se excluyan o sean obsoletos conceptos como virtud o carácter. Sin embargo, el mismo Kant muestra que estos conceptos, tan propios de otras tradiciones éticas, son dependientes e incluso reducibles a la noción más básica y fundamental de deber. Esto es evidente en la *Metafísica de las costumbres* en donde se reduce la noción de virtud a la deontología, o en las *Lecciones de ética* en donde se plantea una lectura deontológica de múltiples fenómenos de orden práctico (p.e. fe, suicidio, deseo sexual, ocio, disciplina, amistad, mentira, entre muchos otros).

El rasgo principal del universo ético es que incluye fenómenos prácticos que no pueden ser capturados por la moralidad. Específicamente, la ética también comporta obligaciones, solo que estas no deben ser confundidas con la obligación moral. Contrastemos, a manera de ejemplo, una obligación moral con una obligación ética: la promesa es una obligación que se adquiere voluntariamente y usualmente se asocia con una obligación moral (aunque es posible también que alguien se sienta obligado a cumplirla por razónes distintas al deber); pero existen muchas otras obligaciones que, en contraste, no se adquieren voluntariamente—son resultado de la fortuna moral, por ejemplo— y surgen dentro de contextos que escapan a nuestro control y que dependen de "una función, una posición o una relación" (Williams, 2006c, p. 7). Por ejemplo, la obligación que siente un hombre de salvar a su esposa que está en peligro no está determinada –incluso, no puede estar determinada– por el deber moral, pero sí por el amor que siente por ella.[92] Aunque la moralidad y la ética supongan obligaciones, se trata de fenómenos diferentes, y lo que es más importante, no es posible reducir la obligación de la ética a la obligación moral. Este punto es central y lo ampliaré más adelante.

Además de la obligación, la ética también comprende otro tipo de fenómenos distintos e irreductibles a la moralidad: (i) *sacrificio*: cuando una persona sacrifica su vida por una causa, su acción no siempre —casi nunca— se explica a la luz de un deber moral, sino a la luz de términos propios de la ética, como: "proyecto", "sentido" o "ideal" (Rivera, 2007). (ii) *Sentido [meaning] de la vida*: cuando una persona se pregunta por aquello que dota de sentido a su vida, es verdad que puede responder a esta pregunta sobre la base de criterios morales (Herman, 1983), sin embargo, la pregunta misma constituye una preocupación de carácter ético (Williams, 2006c). (iii) *Integridad:* si, para poner el caso central de esta investigación, una persona se preocupa genuinamente por su integridad (§4.2), ello solo es posible si ella asume una perspectiva ética de su propia vida. Sugiero estas ideas, repito, para señalar que hay fenómenos que hacen parte de la vida práctica de los seres humanos que no pueden ser capturados por la moralidad. Estos fenómenos, y con seguridad otros más, son los que configuran el universo de la ética.

[92] Williams analiza este ejemplo en "Persons, Character and Morality" (1981d). Allí indica que si el hombre necesitara una justificación moral para salvar a su esposa se le acusaría de pensar de más [*one tought too many*]. El ejemplo es usado por el autor para mostrar que "una persona no necesita, y tal vez no debería, estar pensando en lo que es justificable moralmente todo el tiempo" (Wolf, 2012, p. 71).

Al trazar la distinción entre moralidad y ética se aprecia mejor la diferencia de estos dos universos y, además, se puede intuir en qué medida el reduccionismo deontológico no puede ser exitoso, ya que estos fenómenos éticos que aquí apenas menciono difícilmente puede ser traducidos o reducidos a la obligación moral. Si el reduccionismo deontológico es falso, como espero demostrar más adelante, entonces eso demostraría igualmente el límite de la moralidad. No busco negar la idea de que el concepto de deber sea importante para la evaluación de las acciones. No creo que este sea un camino promisorio, en parte porque el concepto de deber es importante en nuestra comprensión de una vida moralmente guiada. Dudo, pues, que el problema radique en el concepto mismo de deber. El problema, más bien, radica en el reduccionismo de la moralidad; es decir, en la idea de explicar toda la vida práctica a la luz del concepto de deber. Considero que la acción humana, en un sentido normativamente relevante, no se expresa ni puede expresarse siempre en términos de deberes u obligaciones morales.

Si admitimos la distinción ética/moralidad, tal como ha sido presentada acá, la tesis del reduccionismo deontológico se debilita. La razón es obvia: el reduccionismo subsume toda la acción humana normativamente relevante al deber moral, mientras que la ética amplía el universo normativo, al punto de que, además del deber, introduce otras consideraciones y valores prácticos que no dependen ni tampoco se agotan en la moralidad ni en la obligación moral. Hay, pues, otro tipo de consideraciones prácticas que pertenecen al universo de la ética y que apuntan más allá del paradigma de moralidad. Estas consideraciones están legitimadas, pues ellas hacen un contrapeso normativo a la noción de deber moral. Es esto lo que intentaré justificar en las siguientes secciones.

b. El concepto de importancia

La distinción entre ética y moralidad ha sido introducida con el propósito de que sirva de marco de comprensión del terreno en el que se mueve mi crítica a la tesis de la primacía normativa de la moralidad. Lo que he dicho hasta ahora, sin embargo, solo es suficiente para señalar esta distinción, pero hace falta desarrollarla adecuadamente. Para empezar, solo he delimitado negativamente la ética como aquello que no es moral, por ello aún es preciso ofrecer una caracterización positiva del concepto. Espero con esto brindar una justificación filosófica del campo de la ética y ahondar más en la distinción entre ética y moralidad.

En lo que sigue, mi propuesta será caracterizar la ética a través de la explicación de su valor normativo. ¿Qué valor normativo tiene la ética, si es que tiene algún valor? ¿Cuál es la fuente ese valor? Me interesa mostrar que la ética sí tiene un valor normativo que puede hacer de contrapeso a la moralidad. Para justificar esta posición introduciré y caracterizaré una noción normativamente cargada que, como espero mostrar, pertenece al universo ético y que explica la fuerza normativa de criterios no morales. Me refiero a la noción de *importancia*.

Es un hecho —en sentido práctico— que las personas consideramos que hay cosas importantes. La noción de importancia indiscutiblemente existe en la vida práctica de los seres humanos. Existe, por un lado, un sentido relativo de importancia: "alguien encuentra que una cosa determinada le resulta importante [para él o ellos o nosotros]" (Williams, 2016b, p. 224). Por otro lado, junto a esta idea de importancia, también existe la noción *simpliciter* de importancia: algo es importante sin más (*ibid.*). Siguiendo a Williams, el sentido *simpliciter* no implica que haya algo que sea universalmente importante.

Más interesante para mí es el sentido relativo de importancia. He introducido el sentido de que algo es importante sin más para establecer un contraste respecto de la relativa importancia que algo puede llegar a tener, y para entender también que es posible considerar algo importante sin que ello signifique que ese algo sea importante para todos (p.e. me importa –en sentido subjetivo y, por tanto, relativo– articular en esta investigación una crítica al sistema moral kantiano, pero eso tiene sin cuidado o poco le importa a casi todo el resto de la humanidad). También puede suceder que haya cosas importantes *simpliciter* pero que a alguien no le resulten importantes (p.e. el calentamiento global es importante para todos los seres humanos en la medida en que afecta la posibilidad misma de la vida en el planeta, y aun así puede no importarme en lo absoluto, incluso puedo no considerar su existencia). La noción de importancia relativa no es incluyente ni universal. Lo que sí es universal es el hecho de que todos atribuimos importancia a determinadas cosas en nuestras vidas.

A partir del hecho de que hay cosas importantes en nuestras vidas es posible derivar una conclusión teórica muy interesante, y es que lo importante no puede identificarse, traducirse, ni mucho menos reducirse al mismo campo de la pregunta kantiana 'cómo debemos comportarnos'. Aquello que nos importa, sin duda, determina estructuralmente nuestra manera de actuar, pero se trata de una

forma de determinación que no se expresa en términos de la obligación ni el deber moral.

> La [moral]93 se concentra en el problema de ordenar nuestras relaciones con otras personas. Se ocupa especialmente del contraste entre lo correcto y lo incorrecto, y del fundamento y los límites de la obligación moral. Por otro lado, llegamos a [otra] rama de investigación porque nos interesa decidir qué hacer con nosotros mismos y porque, en consecuencia, debemos entender qué es importante o, más bien, qué es importante para nosotros. (Frankfurt, 2006a, p. 119-120; énfasis en el original)

Lo que es importante para nosotros no es lo mismo que aquello que es correcto/incorrecto para nosotros. Se trata de otro aspecto de la acción humana. La normatividad moral por supuesto que exige y determina nuestras acciones, pero se trata de una exigencia que busca que esas acciones se correspondan con el deber moral; en cambio, aquello que es importante para nosotros no necesariamente se adecúa a lo debido o con lo correcto desde el punto de vista moral.

El concepto mismo de importancia, sin embargo, es muy difícil de caracterizar. Encuentro dos razones que justifican esta dificultad. La primera, la extraigo de un comentario de Williams, citado arriba: cualquier explicación del concepto de importancia "necesariamente está afectada por lo que [las personas] consideran importante" (2006b, p. 182). En otras palabras, dado que la noción de importancia es un concepto práctico, cualquier explicación de lo importante supone destacar o priorizar las características prácticas de esta noción, pero esto ya es una forma de usar el concepto mismo de importancia, pues supone distinguir las características prácticas relevantes de aquellas que no lo son. En otras palabras, el solo intento de destacar o de otorgarle prioridad a ciertas características más que a otras es ya un uso de la noción de importancia. De una manera más técnica, si se toma "importancia" como el *explanandum*, cualquier tipo de *explanans* supone este *explanandum* en su formulación. Esta es básicamente la descripción de una petición de principio.

Otra razón que explica la dificultad de definir la noción de importancia tiene que ver con un problema de circularidad. Este argumento es desarrollado por Harry Frankfurt en su ensayo 'La

93 En el original, Frankfurt habla de "ética", no de "moral". Introduzco este cambio para mantener coherencia respecto de la distinción moral-ética; además Frankfurt no diferencia la moralidad de la ética, de modo que en nada se modifica el sentido de la cita.

importancia de lo que nos preocupa' (2006a). Allí se presenta el argumento muy condensado y hasta de un modo críptico. En todo caso, con ánimo expositivo, propongo el siguiente esquema para el argumento:

1. "Es razonable suponer que las cosas tienen importancia sólo en virtud de la influencia que tienen: si el hecho de que algo existiera o de que tuviera ciertas características no tuviera ninguna influencia sobre nada, entonces ni la existencia de esa cosa ni sus características tendrían importancia alguna" (Frankfurt, 2006a, p. 122). La hipótesis, entonces, es que el concepto de importancia podría definirse en términos de la noción de influencia: lo importante tiene influencia, y sin influencia nada es importante.

2. Sin embargo, "todas las cosas tienen *alguna* influencia" (Frankfurt, 2006a, p. 122; énfasis en el original), es decir, todas las cosas comportan directa o indirectamente algún tipo de diferencia para nosotros. Una cosa que no tenga absolutamente ninguna influencia es como si no existiera para nosotros.

3. Si todo tiene alguna influencia, pregunta Frankfurt, "¿cómo es posible que algo [con alguna influencia] en verdad carezca de importancia?" (Frankfurt, 2006a, p. 122). Esta pregunta atiende a la relación entre la importancia de algo y su influencia. Si bien todo tiene alguna influencia, no todo es importante, ya que si todo fuera importante nada lo sería. La importancia de algo y su influencia, por tanto, no son ideas co-extensivas. Es cierto que la influencia de una cosa puede ser importante, pero en otras ocasiones —y esto es lo relevante— "la influencia de esa cosa no tiene importancia en *sí misma*" (*ibid.*; énfasis añadido).

4. Si definiéramos la importancia en términos de influencia (como se intentó en el punto 1), entonces el análisis del concepto de importancia debería incluir la condición de que nada es importante, a menos que la influencia que tenga sea importante. En otras palabras, lo que es importante lo es virtud de que su influencia sea importante. Pero este planteamiento supone una circularidad, pues para explicar lo que es importante habría que usar el concepto de importancia (influencia importante).

Las dificultades para explicar o para definir el concepto de importancia no deslegitiman la noción misma. De hecho, como sugiere Frankfurt, probablemente sea el caso de que esta noción es primitiva; es decir, que parece un concepto tan "fundamental que quizá resulte absolutamente imposible hacer un análisis satisfactorio de él" (2006a, p. 122).

Aun cuando sea difícil establecer una definición del concepto de importancia, es posible establecer una relación entre este y otros fenómenos prácticos que hacen parte de la comprensión de la acción humana. Específicamente, tal y como lo muestra Frankfurt, la importancia se relaciona muy estrechamente con el fenómeno de 'lo que nos preocupa'.[94]

> Si hay algo por lo que una persona sí se preocupa, entonces eso significa que es importante para ella. Esto no se debe a que la preocupación implique algún tipo de juicio infalible con respecto a la importancia de su objeto. Más bien es porque la preocupación por algo vuelve [makes] importante esa cosa para la persona que se preocupa por ella. (Frankfurt, 2006a, p. 134-5)

La anterior corresponde a la formulación del *Principio de preocupación-importancia* (Benbaji, 2001, p. 415). Este principio afirma que la preocupación que una persona tiene por algo determina que esto le resulte importante. Entonces, una característica de aquello que nos importa –no en sentido *simpliciter*, sino en sentido relativo– es que está asociado a lo que nos preocupa. Algo que no nos importe es también algo sobre lo que no recae ninguna preocupación.

No quisiera reducir el concepto de importancia a un fenómeno subjetivo relativo a las preocupaciones singulares de un agente. No se puede subjetivizar este concepto a tal grado que lo importante resulte inescrutable para otras personas. Habría que pensar en una forma de entender

[94] "Desde hace tiempo, los filósofos dedican una atención sistemática a dos amplios conjuntos principales de preguntas, que surgen a raíz de una preocupación con un aspecto omnipresente, imperioso y problemático de nuestras vidas. En el primer conjunto, que constituye el ámbito de la epistemología, las preguntas derivan de una u otra manera de nuestro interés en decidir *qué creer*. El tema general de las preguntas del segundo conjunto es *cómo comportarnos*, es decir, el contenido de la [moral]. También es posible delinear una tercera rama de investigación, que tiene que ver con una serie de cuestiones que pertenecen a otra área temática y son un objeto de interés fundamental de la existencia humana, es decir, qué debe preocuparnos". (Frankfurt, 2006a, p. 119-120; énfasis en el original).

lo importante sin reducirlo a un estado psicológico. Y con todo, no se puede tampoco perder de vista que lo importante tiene la característica de preocuparle a alguien. Si se pierde de vista esto último, no habría ningún vínculo psicológico que explique el interés de una persona por eso que encuentra importante. Lo importante, entonces, es relativo siempre a un agente. No es posible explicar la preocupación que alguien tiene por algo sin apelar a su conjunto motivacional, a su carácter, al tipo de persona específica que es ese alguien. Incluso si algo fuera importante de modo universal, lo sería solo porque le preocupa a todas las personas, no porque sea intrínsecamente importante. Con esto sugiero que la noción de importancia se explicaría mejor a través de una concepción internalista de la acción humana, pero no me detendré a justificar esta afirmación. Mi punto simplemente es que no se puede perder de vista la fuerza psicológica subjetiva de lo que nos importa.

c. Prioridad deliberativa

Lo *importante*, a pesar de ser una noción difícil de precisar, cuenta con una innegable fuerza normativa. Prueba de ello es que tiene peso en nuestros procesos de deliberación. Muchas veces, cuando deliberamos, optamos por aquella opción que es más importante para nosotros. Esta forma de *prioridad deliberativa*, como la llama Williams (2016, p. 225), solo es posible porque lo importante tiene fuerza normativa. Sin embargo, como lo señala también Williams, no se debe confundir la importancia con la prioridad deliberativa. "Una consideración tiene gran prioridad deliberativa para nosotros si le damos más peso que a las demás consideraciones en nuestras deliberaciones" (*ibid.*). Tanto la prioridad deliberativa como lo importante marcan una especie de jerarquía: la prioridad deliberativa se refiere a lo que "tiene más peso" en la deliberación; lo importante determina aquello que tiene mayor valor para alguien. Además, la prioridad deliberativa, al igual que lo importante, también es relativa a una persona o a un grupo de personas, es decir, aquello que es importante o aquello que tiene prioridad deliberativa puede variar dependiendo de la persona o el contexto social.

Sin embargo, no es correcto confundir la prioridad deliberativa con lo importante: se puede pensar que si algo es importante, entonces posee prioridad deliberativa, y esta idea es correcta hasta cierto punto; pero, insisto, esto no puede llevarnos a identificar ambos conceptos: lo que es importante

para alguien puede ser algo sobre lo que no podemos hacer nada (Williams, 2016b, p. 225), algo que escapa a nuestra agencia y, por tanto, algo sobre lo que no tiene sentido deliberar[95]. Con esto quiero señalar que lo que nos resulta importante no siempre es objeto de deliberación (no siempre podemos elegir lo que es importante para nosotros), aunque sí que incluimos lo que es importante para nosotros como parte de nuestras deliberaciones (deliberamos usando como criterio aquello que es importante).

Aun cuando la prioridad deliberativa y lo importante no sean lo mismo, tienen un aspecto en común que es esencial: a través de ellos se captura la idea de que diferentes cosas pueden ser importantes para alguien y también diferentes consideraciones pueden tener prioridad deliberativa. Esta puede parecer una idea trivial, pero tiene implicaciones significativas. No existe una única cosa o consideración que tenga prioridad en la deliberación y tampoco existe una única cosa que sea importante. En otras palabras, ni la prioridad deliberativa ni lo importante imponen restricciones a los objetos que pueden ser considerados en la deliberación o que pueden hacer parte de lo que nos importa. Existen diferentes objetos o consideraciones que pueden tener prioridad deliberativa o ser importantes.

La relevancia filosófica de lo anterior es que no se puede decir que las consideraciones morales sean las únicas que tienen prioridad deliberativa. En efecto, si se tratara de delimitar la noción de prioridad deliberativa a un solo objeto se perdería el sentido de esta noción: justamente el espíritu de la prioridad deliberativa es establecer una jerarquía entre distintos objetos (consideraciones, criterios, principios, máximas, deberes), así que antes de proclamar la prioridad deliberativa de la moralidad (deberes y obligaciones morales) es necesario preguntarse previamente por qué la moralidad tiene prioridad frente a otras consideraciones no necesariamente morales. Por lo tanto, no se puede asumir, sin justificación, que la moral tenga prioridad deliberativa. Una consideración específicamente moral solo puede tener prioridad deliberativa cuando se le otorga tal prioridad, pero este es un derecho adquirido tras la deliberación que no se puede ser asumido *a priori* o sin

[95] Esta idea es muy cercana al *motto* aristotélico "se delibera sobre los medios, no sobre los fines" (*EN* 1113a). Sin ánimo de trazar una identidad entre las nociones de fin y el de importancia (aunque es verdad que ambos conceptos guían nuestras acciones), lo interesante es que en los dos casos se postula un marco de sentido para la deliberación, pero la deliberación misma no puede usar este marco como su objeto de deliberación.

justificación alguna.

Todo esto significa que la pregunta por el objeto de la prioridad deliberativa es una cuestión abierta a distintos objetos. "[L]a prioridad deliberativa puede relacionar consideraciones de diverso tipo" (Williams, 2016b, p. 226). Y es a partir de la pluralidad de consideraciones morales y no-morales que un agente establece qué tiene prioridad deliberativa. Una vez que se ha establecido qué objeto o consideración "pesa más en la deliberación", es posible guiar los juicios prácticos. Si, como el rigorista, se asume que el deber moral tiene siempre prioridad sobre otras consideraciones, debemos cumplir incondicionalmente el deber. Si, como Williams, se cuestiona la prioridad deliberativa de la moralidad, entonces hay otras consideraciones, como la integridad, que pueden tener más prioridad deliberativa que la moralidad.

Dado que la prioridad deliberativa es neutra con respecto de los objetos que pueden ser materia de deliberación, y dada la pluralidad de objetos posibles en la deliberación, se impone la pregunta de qué clase de consideraciones tienen más prioridad que otras. Poder formular esta pregunta es ya una victoria en contra de la prioridad normativa de la moralidad, pues la prioridad deliberativa sitúa las consideraciones morales y no-morales al mismo nivel. Para los intereses de mi crítica a la tesis de la prioridad normativa de la moralidad, poder plantear la pregunta por el objeto de la prioridad deliberativa implica que la moralidad ya no se impone necesariamente sobre otro tipo de consideraciones de carácter práctico. Evidentemente, con la noción de prioridad se mantiene una jerarquía de valores, pero esta jerarquía no siempre está determinada por la estructura de la moralidad. Las consideraciones morales son unas, entre otras.

d. Necesidad práctica

He dicho antes que la noción de importancia no puede identificarse con la prioridad deliberativa. Sin embargo, ambas nociones, cuando coinciden, dan lugar a un fenómeno práctico muy interesante:

> Una conclusión deliberativa que incorpora una consideración que tiene tanto la mayor prioridad deliberativa, como la mayor importancia (por lo menos para el agente) puede adoptar una forma especial y convertirse en la conclusión no solo de que uno ha de hacer

> algo en concreto, sino que tiene que [must] hacerlo y de que no puede hacer otra cosa. Podemos llamar a esto una conclusión debida a la necesidad práctica (Williams, 2016b, p. 230; énfasis en el original).

La necesidad práctica está asociada a la idea de que alguien, cuando delibera, concluye que tiene que realizar cierta acción (Williams, 1993a, p. 159). La necesidad se refiere a este *tener que,* que una persona asocia a una obligación y que es tan fuerte que ella no puede hacer otra cosa. Esta necesidad es un tipo especial de obligación a la que el agente no puede más que responder. El caso de Paul Gauguin presentado por Williams ilustra este tipo de necesidad (ver §4.2.1c): su vocación lo "obligaba" a abandonar a su familia para pintar en Tahití. Williams también presenta el caso del Áyax de Sófocles cuyo sentido de honor y virtud lo conducen al suicidio: "Yo voy allí donde debo encaminarme" (Áyax, 690). Harry Frankfurt, por su parte, recurre a la frase de Lutero para ilustrar este fenómeno:[96] "aquí estoy, no puedo hacer otra cosa"[97].

Aun cuando no sea evidente en nuestro vocabulario o no tengamos consciencia de su influencia, el fenómeno de la necesidad práctica no es ajeno a nuestra vida. La principal dificultad para dar cuenta del fenómeno es que ha sido confundido o reducido a otro fenómeno práctico que también tiene que ver con la necesidad pero que, sin embargo, es distinto: la obligación moral, en el sentido kantiano del término.

Kathi Beier (2017), en un intento por caracterizar la noción de necesidad práctica y por distinguirla de la obligación moral kantiana, ha rastreado la ascendencia de esta noción hasta Aristóteles. Su trabajo puede leerse como un esfuerzo por desligar la necesidad práctica de la tradición kantiana. Según la autora, este fenómeno práctico puede encontrarse en el modelo de deliberación

---

[96] La frase de Lutero ha servido en la literatura filosófica como la formulación más sintética de la necesidad práctica (ver el compendio de Bauer *et al.*, 2017). Esta frase expresa la idea de que, ante la necesidad práctica, una persona se ve a sí misma sin otra alternativa. Sin embargo, esta la frase no captura completamente el elemento de identificación que me parece central en el concepto de necesidad práctica.

[97] Como bien lo expresa Phillippa Foot, la noción de necesidad práctica es un "pensamiento fugitivo" [*fugitive thought*] (1972, p. 311) y es difícil capturarlo con precisión. Existen otros casos más cotidianos de la necesidad práctica. Uno de ellos es la experiencia del amor. Frankfurt (2006c; 2007) ha desarrollado principalmente esta línea de pensamiento a través de su idea de necesidad volitiva (sobre este punto ver Schechtman, 2017). Otro caso de necesidad práctica tiene que ver con el perdón: en ocasiones una acción resulta imperdonable para la víctima, ella simplemente no puede perdonar a su agresor (ver Hallich, 2017).

aristotélico[98]. Para el propósito que persigo, el argumento de Beier no es tan interesante, como el espíritu de su análisis: explicar la necesidad práctica a partir de un modelo no kantiano y con ello desvincular la idea de la necesidad práctica de la noción de obligación moral.

Un rasgo del análisis aristotélico de Beier, que también está presente en Frankfurt (2007) y Williams (1993a), y con el que me comprometo, es la idea de que la necesidad práctica no puede desligarse del carácter. Como bien lo expresa Williams, las necesidades prácticas no solo "establecen límites para el carácter y proporcionan condiciones para él, sino también, en parte, constituyen su substancia" (1993a, p. 130). Esta misma idea en palabras de Beier se expresa de la siguiente manera: "Las acciones realizadas de acuerdo con la necesidad práctica son acciones que expresan un fuerte sentido de identidad personal. […] Que el agente desarrolle un fuerte sentido de identidad personal significa que hay algo que es absolutamente importante para él […]" (Beier, 2017, p. 136).

El término *necesidad práctica* es, entonces, una expresión del carácter y de la identidad de una persona. Si Baier, Frankfurt y Williams tienen razón —y creo que la tienen— entonces aquello que la persona 'tiene que hacer' es una manifestación del tipo de persona que ella es. Y al mismo tiempo, sin embargo, ese *tener que hacer* es algo ante lo que no hay alternativa y, por lo tanto, parece estar más allá del control esa persona. Así caracterizada, se puede entrever una tensión en el concepto mismo de necesidad práctica. Esta tensión se genera por el choque de dos intuiciones. La primera es que el carácter es algo que en un sentido relevante está bajo nuestro control y

[98] Beier extrae su conclusión del análisis de EN III. Aristóteles allí discute la noción de acción involuntaria. "Parece, pues, que cosas involuntarias son las que se hacen por fuerza o por ignorancia" (*EN* 1110a1). Esta fuerza, dice Aristóteles, ha de entenderse como un "principio externo" (EN 1110a2–4), como en el ejemplo del viento que nos empuja o cuando somos movidos por hombres que nos tienen en su poder. En cambio, dice Beier, la acción voluntaria puede estar sometida a una necesidad dada por una causa interna (Beier, 2017, p. 137). Esta necesidad se explica a través de la deliberación práctica: "una acción es necesaria si no podemos alcanzar el fin –considerado bueno– sin ella" (*ibid.*). Por ejemplo, es necesario tomar la medicina para curar la enfermedad, es necesario entrenarse para ganar en las olimpiadas. No me comprometo con esta conclusión de Beier, y aunque su análisis me parece muy prometedor, considero que éste deriva en una noción reducida de necesidad, pues se trata de una necesidad meramente instrumental: cualquier acción que sea un medio para un fin resulta necesario. Esta idea deja de lado la conexión que existe entre necesidad práctica y los fines (en mi lectura, es importante la idea de fines porque ellos se relacionan con los proyectos fundamentales y con la razón que tenga una persona para vivir). La carencia en el análisis de Beier se explica porque en el modelo aristotélico los fines no son objeto de deliberación.

depende de nosotros (de otra manera no podríamos atribuirnos a nosotros mismos responsabilidad por quiénes somos). La segunda intuición es que la obligación que impone la necesidad práctica es algo frente a lo que no podemos hacer otra cosa. Si se aceptan ambas intuiciones, cómo es posible entonces que la fuente de la necesidad sea nuestro propio carácter, o en otras palabras, ¿cómo es posible que algo que depende de nosotros dé lugar a obligaciones frente a las que no podemos hacer otra cosa? La tensión, entonces, consiste en el hecho de que para alguien puede no existir una alternativa frente a la necesidad práctica, aun cuando esta necesidad sea una expresión de nuestra identidad y de nuestro carácter.

Esta tensión se puede expresar conceptualmente de la siguiente manera: por un lado, la necesidad práctica contiene un elemento de identificación: el agente se identifica con aquello que le impone la necesidad, pues ella es una expresión de su propio carácter. Por otro lado, la necesidad práctica contiene un aspecto de necesidad: el agente tiene que hacer lo que lo obliga la necesidad práctica, y no puede hacer otra cosa.[99] El concepto de necesidad práctica engloba entonces estas dos caras que parecen ser tan antagónicas. ¿Cómo puede ser que aquello que es significativamente expresivo de nosotros mismos sea al mismo tiempo la fuente de exigencias frente a las que no tenemos alternativa?

La tensión en torno al concepto de necesidad práctica, sin embargo, no significa que este sea un concepto contradictorio o imposible. Como dije, la tensión surge por el choque de las dos intuiciones que caractericé arriba. Al cuestionar alguna de estas intuiciones la tensión podría desaparecer. Por ejemplo, el primer supuesto de que el carácter está enteramente en nuestro control se puede cuestionar a través del fenómeno de la fortuna moral (Williams, 1981c; Rosell, 2009). Y en cuanto a la segunda intuición, el hecho de que una necesidad práctica no nos deje otra alternativa para actuar no significa que el agente sienta que ha perdido el control cuando actúa de acuerdo con esta necesidad (Frankfurt, 2006c; 2007). Ambas intuiciones, por tanto, son muy

[99] Estas características, en conjunto, permiten distinguir, al menos preliminarmente, la necesidad práctica de otros fenómenos prácticos. La necesidad práctica no puede confundirse con la coacción porque aquella no es una necesidad externa. Tampoco puede identificarse con la compulsión o adicción porque, siguiendo la terminología de Frankfurt, la persona se identifica con aquello que impone la necesidad práctica (2007, p. 216). Y tampoco puede decirse que la necesidad práctica sea una pasión o compulsión emocional porque la necesidad es estable, no es contingente, debido a que es una expresión del carácter.

cuestionables. Y justamente, la noción de necesidad práctica, más que adecuarse a ellas, arroja dudas sobre estas intuiciones que causan la apartente tensión. El concepto de necesidad práctica, más bien, propone pensar el carácter como algo compatible con la necesidad. O extrapolando esta idea, la necesidad práctica muestra que la agencia humana, por un lado, y la necesidad, por otro, no son incompatibles (*cf.* Beier, 2017, p. 136).

Dejando de lado esta discusión, pasemos ahora a caracterizar específicamente la necesidad práctica. Para ello trazaré un contraste entre la necesidad práctica y un caso de deliberación en el que ella no esté involucrada.

> Si A quiere X, y si es cierto que si quiere X debe hacer Y, de ahí no se desprende que tenga que hacer Y; eso solamente se concluiría si, además, X es lo que tiene que intentar lograr. Así, en primera persona: si concluyo que tengo que hacer Y, entonces es porque me he dado cuenta no solamente de que es el único medio para un objetivo que tengo, sino de que es lo único que puedo hacer. (Williams, 1993a, p. 161)

Esta sería la descripción del tipo de deliberación que involucra la necesidad práctica y que contrasta con el tipo de deliberación que usualmente realizamos. En las deliberaciones más comunes, no existe una única cosa que el agente pueda hacer, rara vez el resto de alternativas son imposibles, y justamente lo que caracteriza la necesidad práctica es que las otras opciones son "literalmente imposibles" (Williams, 1993a, p. 161). Lo que usualmente sucede en la deliberación donde no hay necesidad práctica es que las otras opciones son excluidas porque X es la opción para la que hay mejores razones.

Lo anterior apunta a la diferencia esencial entre la deliberación de la necesidad práctica y cualquier otra forma de deliberación exitosa: la necesidad práctica no se ajusta al modelo de las "mejores razones". Si este fuera el caso, la necesidad práctica no sería muy distinta de una deliberación cualquiera en la que la opción X tiene a su favor las mejores razones. Una deliberación cualquiera que concluya que *debo* hacer Y (porque si quiero X debo hacer Y) no necesariamente es un caso de necesidad práctica. El *tener que* de la necesidad práctica es distinto.

La deliberación que involucra la necesidad práctica no tiene que ser necesariamente certera ni

firme. Pensemos un escenario de deliberación en el que una de las opciones disponibles tiene las mejores razones a su favor y las tiene de un modo contundente. El peso de estas razones sobrepasa por mucho las razones de las otras opciones. En este escenario, la persona sabe lo que tiene que hacer. En esta deliberación, sin embargo, no es muy clara la diferencia entre la necesidad práctica y cualquier otra deliberación exitosa, excepto porque las razones para X son evidentemente mejores. En casos como este, según Williams, la necesidad práctica se confunde con la certeza y con la firmeza, pero estas cualidades no son esenciales a la necesidad práctica (Williams, 1993a, p. 161). En un caso de deliberación exitosa, las otras opciones son imposibles de modo metafórico, pues están disponibles para el agente, solo que las razones para estas otras opciones son muy débiles o representan limitaciones muy costosas.

En cambio, una deliberación que involucre la necesidad práctica puede desprenderse "solamente después de una larga y anhelosa consideración de las alternativas, y puede albergar tal creencia, aun si se está todavía inseguro al respecto y percibe aun claramente los poderosos méritos de los cursos alternativos" (Williams, 1993a, p. 161). Piénsese en el escenario del soldado de Sartre (2017): no sabe si ir a la guerra o quedarse cuidando a su madre. Tiene muy buenas razones para llevar a cabo los dos cursos de acción; y aun así, si su deber patriota es una necesidad práctica para él, tendrá que ir a la guerra a pesar del bienestar de su madre.

La diferencia entre el escenario en el que la deliberación es firme y el escenario en el que la deliberación es "insegura" es que en el primero la persona da por sentado los objetivos o limitaciones que tiene y eso la lleva a preferir una opción sobre las demás. En cambio, en el segundo escenario, el del soldado, la necesidad afecta a los propios objetivos y a las limitaciones mismas.

Esta comparación entre distintos escenarios de deliberación sirve para mostrar que la necesidad práctica no excluye la agencia, pues el agente delibera sobre aquello que es objeto de su necesidad práctica. Y con todo, incluso tras deliberar no implica que él pueda hacer otra cosa. La deliberación solo confirma lo que el agente tiene que hacer, o confirma que no puede hacer otra cosa.

La necesidad práctica impone a alguien límites respecto de aquello que puede hacer. Si tiene que hacer algo, y si las otras opciones son literalmente imposibles, es porque la necesidad "acarrea

consigo una noción correspondiente de imposibilidad" (Williams, 1993a, p. 162). La imposibilidad es la otra cara de la necesidad práctica: *el tengo que,* evidentemente, sugiere un no-puedo (*cf.* Altshuler, 2013, p. 42).

Supongamos un caso en que están disponibles las opciones A, B y C, y la necesidad práctica me lleva a hacer A. Luego B y C serán opciones que no puedo hacer, imposibilidades para mí. Insisto, esta imposibilidad es metafórica, porque son opciones que está disponibles, solo que A sobrepasa por mucho a las opciones B y C. Lo importante de este caso es que la imposibilidad se deriva de la necesidad práctica. En otro escenario, en cambio, es posible que la imposibilidad misma tenga prioridad en la deliberación. B y C serían cosas que no puedo concebir hacer y, por lo tanto, son imposibles para mí; de manera que A es lo que tengo que hacer. En este segundo caso, la necesidad se deriva de la imposibilidad. No es fácil entender qué explica qué, o qué noción es más básica entre la necesidad y la imposibilidad. Lo cierto, en todo caso, es que la necesidad práctica se corresponde con la noción de imposibilidad.

Ahora bien, la imposibilidad relevante aquí no lo es en un sentido metafísico, ni lógico, ni físico. La imposibilidad que surge de la necesidad práctica es, naturalmente, una imposibilidad de orden práctico, y más precisamente, se trata de una incapacidad vista desde el punto de vista de la primera persona que solo le atañe al agente mismo: "lo que reconozco cuando concluyo en la deliberación que no puedo hacer cierta cosa es cierta incapacidad *mía*" (Williams, 1993a, p. 164; énfasis añadido). No todos los agentes son iguales y, por lo tanto, no todos tienen las mismas limitaciones. Las opciones disponibles para un agente pueden ser concebibles por él mismo o por un observador externo, pero el punto es que el agente, desde su perspectiva, no puede considerarlas como opciones serias. La única opción seria es aquella por la que tiene que optar, el resto (incluso si tienen peso) resultan excluidas; a la luz de la necesidad práctica, estas opciones son imposibilidades.

También puede suceder que las imposibilidades, aunque sean concebibles, se excluyen de entrada porque suponen una gran carga moral o prudencial, o porque no existen los medios materiales para llevar la acción a cabo o porque no es una posibilidad física. El punto es que estas imposibilidades son impensables, al menos, nuevamente, en un sentido metafórico. Si bien ellas alcanzan el umbral de la deliberación, el agente no piensa en ellas como opciones realizables.

Por otra parte, una imposibilidad es literalmente impensable cuando, durante la deliberación, un agente no es consciente de una imposibilidad que un observador externo, sin embargo, sí puede registrar. Cuando esto sucede, la imposibilidad resulta impensable en el sentido estricto y literal del término. Se trata de un tipo especial de imposibilidad que ni siquiera es concebible por el agente.

Esta clasificación de las distintinas imposibilidades prácticas (concebibles o impensables) podrían calificarse todas ellas como *incapacidades del carácter* (Williams, 1993a, p. 165). Estas incapacidades demarcan los límites entre lo que uno está dispuesto a hacer y lo que simplemente no puede hacer. De manera que las necesidades prácticas, junto con las imposibilidades prácticas, "no solamente pueden establecer límites para el carácter y proporcionar condiciones para él, sino también, en parte, constituir su substancia" (p. 166).

La necesidad práctica configura y delimita el carácter individual de una persona. Su relación con el carácter es tan estrecha que la necesidad expresa los compromisos más propios de la identidad y de la individualidad de una persona: "el hecho de ser una expresión del carácter es quizá la forma más sustancial en que una acción [promovida por la necesidad práctica] puede ser propia" (Williams, 1993a, p. 166). Las acciones u omisiones promovidas por la necesidad práctica son tal vez los indicativos más honestos y reveladores de quién es la persona.

La necesidad práctica, a la vez que fija el curso de acción que alguien emprende, también fija un límite sobre lo que puede hacer ese alguien. Tomar consciencia de este contorno de posibilidades e imposibilidades representa, en palabras de Williams, "hacer un descubrimiento sobre uno mismo" (Williams, 1993a, p. 166). Esto disipa la supuesta tensión interna de la necesidad práctica que describí arriba. Señalé que esta tensión surge por el choque de dos intuiciones: que el carácter depende de uno mismo y que la necesidad práctica impone una obligación que no depende de nosotros. La tensión se disuelve entonces al rechazar la primer intuición, esto es, la idea de que el carácter depende enteramente de uno mismo. En contra de esta idea, no podemos elegir cuáles sean nuestras necesidades prácticas, no al menos el sentido más común del término, y, por lo tanto, no podemos elegir completamente el tipo de acciones que son significativamente propias de nosotros mismos. Es cierto que las necesidades prácticas pueden ser objeto de deliberación, y que

la conclusión de esta deliberación es algo del tipo "tengo que hacer X", pero esta conclusión puede verse la mayoría de veces como un auto-descubrimiento que hace alguien sobre la capacidad o incapacidad de su propio carácter: el auto-descubrimiento consiste en reconocer el alcance de las propias acciones (lo que tienen que hacer) o, también, en reconocer el límite de lo que es incapaz de hacer (lo que no puede hacer).

Todo este análisis del concepto de necesidad práctica tiene importantes repercusiones para el asunto de la integridad personal. En el capítulo 4, traté de interpretar la integridad en términos del carácter de una persona. El carácter incluía, entre otras cosas, los proyectos fundamentales y las razones que una persona tiene para vivir. En este capítulo incluyo la noción de necesidad práctica como otro elemento que enriquece este ya complejo fenómeno del carácter. Carácter, proyectos fundamentales, necesidades prácticas son conceptos que se relacionan estrechamente entre sí. Todos ellos moldean nuestro sentido de identidad y perfilan el tipo de persona que somos.

La relación entre carácter y necesidad práctica fortalece y replantea el problema de la integridad personal. En §4.2, busqué aclarar el sentido en el que el rigorismo moral suponía una amenaza a la integridad del agente. El rigorismo kantiano y también su versión más débil, la tesis de la prevalencia, asumen una jerarquía establecida que otorga prioridad normativa al deber moral. Pero esta jerarquía pasa por alto el fenómeno de la necesidad práctica y su papel en la formación del carácter; y al hacer esto, ella supone una amenaza a la integridad personal. Sin embargo, con la noción de necesidad práctica toma fuerza nuevamente el asunto de la integridad personal: una forma de explicar que alguien no realiza cierta acción porque supone un riesgo para su integridad es constatando que para ella esta acción representa una incapacidad desde el punto de vista de su carácter. Por lo tanto, la integridad de una persona encuentra sus límites en aquello que la necesidad práctica determina como imposible o impensable.

Justamente a la luz de lo anterior, el caso del asesino en la puerta (ver cap. 1) puede ser reinterpretado. Para el rigorista moral kantiano, el deber moral lo insta a ser veraz sin importar las consecuencias para la persona perseguida (ver capítulo 2). En contraste, para un agente no kantiano —un agente genuinamente preocupado por su integridad— lo que él tiene que hacer no necesariamente se corresponde con el deber moral. La persona que enfrenta al asesino puede estar

coaccionada por este, probablemente tema las represalias, y con todo, ella descubre que es el tipo de persona que es incapaz de delatar a su amigo. Le resulta imposible ser veraz con el asesino. Esta imposibilidad es uno de los límites que configura su carácter.

e. Obligación moral y no-moral

La tesis de la prevalencia, como vimos, enunciaba que aunque la moralidad no es incondicional, sí es el criterio práctico más importante en la deliberación. Esta tesis, como he dicho, no elimina la posibilidad de que existan otras fuentes de valor distintas a la moralidad, pero frente a un conflicto práctico, la moralidad tiene prioridad normativa y debe imponerse sobre el resto de criterios.

Con la idea de necesidad práctica he querido criticar la tesis de la prevalencia. La intuición básica es que con el fenómeno de la necesidad práctica se va más allá de la moral kantiana: la noción de obligación moral no es la única especie de obligación a la que estamos atados. "Es obvio que este deber [de la necesidad práctica] no guarda ninguna relación especial con la obligación moral. La pregunta '¿Qué debo hacer?' puede plantearse y responderse sin que en la situación participe ninguna cuestión de obligación moral" (Williams, 1993a, p. 160).

He caracterizado la necesidad práctica como una obligación personal, pues la necesidad que una persona tiene no ha de ser necesaria para alguien más. No podemos exigir que otras personas actúen de acuerdo con nuestras propias necesidades prácticas. (Williams, 2016b, p. 231). Una necesidad práctica no señala una obligación para otros, ni mucho menos señala una obligación moral. Necesidades prácticas que surjan de una vocación, del carácter o de un proyecto fundamental, son traducidas a acciones que están más allá de la moralidad, y por lo tanto, no son obligatorias en el mismo sentido en que lo son los deberes morales: ni se nos puede exigir que las cumplamos, ni es censurable que no se realicen (*cf. ibid.*).

> [E]l agente que realiza una acción semejante puede tener la sensación de que tiene que hacerlo, que no tiene alternativa, y al mismo tiempo reconocer que no podría exigírselo a los demás. Si bien puede pensar que es una exigencia para él, no para los demás, porque

> él es diferente de los demás, normalmente esa diferencia radicará en el hecho de que es justamente él quien tiene esa convicción. Sus sentimientos, y las expectativas sobre qué sentimientos experimentará si no actúa de ese modo, pueden asemejarse mucho a los asociados a las obligaciones. (Williams, 2016b, p. 231-2)

La idea que busco expresar es que la obligación de la necesidad práctica puede ser tan incondicional como la obligación moral. No hay una razón *a priori* por la que la obligación moral deba imponerse por encima de la obligación no-moral. Para justificar esto, sin embargo, habría que decir algo más sobre la obligación moral kantiana.

La filosofía kantiana no es ajena a la noción de necesidad práctica. Ocurre, sin embargo que su interpretación difiere radicalmente de la que he intentado articular aquí. La filosofía kantiana, como ya he dicho, no tiene en cuenta la distinción entre ética y moralidad, y además reduce el fenómeno de la necesidad práctica a la moralidad. En este sentido, la necesidad práctica se reduce únicamente al asunto de la obligación moral. Si fuéramos agentes kantianos, la obligación moral impondría sobre nosotros una acción que tenemos que hacer. El imperativo categórico sería la expresión de una necesidad práctica incondicional llevada hasta sus consecuencias rigoristas.

Kant, además, consideraría que esta forma de ser de la necesidad práctica está dada independientemente de toda inclinación y de todo deseo e inclinación (Williams, 2016b, p. 232). La ley moral obliga al margen de cualquier "acicate empírico". Esto solo es posible, como mostré en §3.4, si se sume un modelo de accón en el yo racional es distinto al yo empírico, y en el que el yo racional está libre de la causalidad mecánica y natural, y por lo tanto, su acción es motivada únicamente por la facultad pura de la razón. La moralidad descansa entonces en la razón pura y nuestra posesión de esta facultad "constituye la parte más profunda y valiosa de nosotros. Un conflicto entre la moralidad y algún otro valor es para Kant un conflicto entre la propia lealtad a la racionalidad y la lealtad a cualquier otra cosa" (Wolf, 1997, p. 312). Ir en contra de la moralidad sería para Kant una forma de traicionar nuestro verdadero y profundo yo, sería "romper la fe en la propia naturaleza racional" (*ibid.*). El punto de Kant, entonces, es que no podría haber un valor o una parte de nosotros que sea más valiosa que el yo racional. No hay nada más allá de la moral por lo que valga la pena renunciar al verdadero yo, "nuestra naturaleza racional es más valiosa y más esencial a nuestra identidad que cualquier otra parte de nosotros" (*ibid.*).

El reconocimiento de la importancia y el valor de nuestra naturaleza racional lo expresa Kant a través del sentimiento de respeto [*Achtung*] ante la ley moral. El respeto hacia la ley es el reconocimiento de la subordinación de la voluntad ante la incondicionalidad del deber moral (*F* 4:400). Esta es básicamente la descripción, en términos kantianos, de la necesidad práctica que impone el deber moral sobre nosotros. Es cierto, entonces, que la filosofía moral kantiana tiene un lugar para el fenómeno de la necesidad práctica; sin embargo, su versión de este fenómeno indica que aquello que tenemos que hacer (o aquello que no podemos hacer) solo puede estar determinado por la obligación moral.

Si este análisis es correcto, para un filósofo rigorista como Kant y, en general, para la tradición kantiana, la noción de obligación moral tiene así la más alta prioridad deliberativa y es además el valor más importante de la vida racional humana. De ahí que el deber moral se imponga sobre cualquier otra consideración práctica. "La más pura de las visiones kantianas ubica la significación de la moralidad en el contexto de la importancia de la motivación moral misma. Lo que más importa es que a las consideraciones morales se les dé la más alta prioridad deliberativa" (Williams, 2016b, p. 226). De ahí que ningún proyecto que entre en conflicto con la moralidad puede ser tan valioso o tan importante como el proyecto moral.

Desde la perspectiva kantiana, entonces, la necesidad práctica no-moral no tiene sentido, pues no tiene la importancia ni la prioridad deliberativa que tiene el deber moral. Sobre esta idea está soportada la tesis de la prioridad normativa de la moralidad.

¿Cuál es, entonces, la crítica a la tesis de la prevalencia? El primer paso en esta dirección ha sido trazar la distinción entre ética y moralidad, que tiene el propósito de mostrar que la moralidad no agota todo lo que incumbe a la vida práctica de los seres humanos (§6.2.2a). Esto significa que el sistema moral kantiano sobrestima el papel del deber en la vida humana y le atribuye la máxima importancia y también la mayor prioridad deliberativa, de ahí que este sistema identifique la necesidad práctica con el deber y la obligación moral.

Con todo, sería un error afirmar que el sistema kantiano desatiende por completo al campo de la ética. De hecho, este sistema moral tiene un respuesta muy clara y categórica a la pregunta socrática: la moralidad kantiana considera que la vida buena es aquella que se corresponde con la

mejor parte de nosotros (la razón), y por tanto la vida buena es aquella que se conduce de acuerdo con el deber moral. Incluso la posición kantiana ofrece una explicación para identificar la vida buena con la vida moral: la vida moral es la única que responde a la mejor parte de nosotros mismos –la parte racional–, de modo que vivir de acuerdo con ella significa vivir la mejor vida humana posible.

El problema de esta respuesta, a mi manera de ver, es que ella denota un menosprecio por la ética: el sistema moral kantiano le resta valor normativo a las consideraciones éticas. Por esta razón, se olvida que la pregunta '¿cómo se ha de vivir?' es una pregunta normativa. En efecto, la pregunta socrática investiga qué cuenta como una buena vida, lo cual supone que habrá que discrimiar entre distintos modos de vida, y seguramente habrá algunos mejores que otros, o eso al menos es lo que intenta investigar la pregunta. De allí que se espere que la respuesta sea de naturaleza normativa. El problema con de Kant —y con el kantismo— consiste en confundir la normatividad de la pregunta socrática con la normatividad moral, pues lo que cuenta como una buena vida es también —y al mismo tiempo— lo que cuenta como criterio para determinar el carácter bueno moralmente de una acción.

La fusión normativa entre ética y moralidad pasa equivocadamente de la idea de que la vida moral es la mejor vida posible a la idea de que los criterios morales son los únicos que tienen importancia para guiar la vida. Pero si la moralidad pretende responder a la pregunta por la vida buena, tendría que entrar en el paradigma de la ética, y desde allí, tendría que asumir la independencia normativa de la ética. Tan pronto la moralidad empieza a ser parte de este universo ético, y este es el punto esencial, es indispensable contar con nuevos parámetros normativos, parámetros que no pueden ser morales. La filosofía kantiana no puede justificar que la vida buena es una vida moral asumiendo la premisa normativa de que la moralidad es lo más importante para la vida humana. No se puede evaluar la respuesta a la pregunta socrática únicamente con los criterios deontológicos de la moralidad, pues, evidentemente, se incurriría con ello en una petición de principio. Mi sugerencia es que algunos –tal vez los más importantes– de estos parámetros normativos para realizar esta evaluación en el plano ético son, sin ánimo exhaustivo, las nociones de importancia y de prioridad deliberativa. Dicho de otra manera, para demostrar que la moralidad es el valor más importante de la vida humana, sin caer en una petición de principio, el kantismo ha de apelar a

criterios éticos (lo importante, la prioridad deliberativa) que se escapan de la normatividad moral kantiana.

Cuando se sitúa la moralidad kantiana en el plano ético, la pregunta relevante ya no es "cómo debo actuar", sino qué tan significativo para la vida buena puede resultar actuar según máximas universalizables. O puesto en otros términos, , ¿qué significa vivir de acuerdo con la filosofía kantiana? ¿Por qué el deber moral es tan importante para una vida buena? Cuando se abordan estas preguntas y se tienen a la mano las nociones de importancia y prioridad deliberativa, se constanta que la vida ética no impone una exclusividad ni una exclusión de otras consideraciones prácticas. La vida ética, dada su preocupación por la pregunta socrática, no puede descartar *a priori* varias formas de vida buena; ella tiene que evaluarlas y mostrar cuáles de ellas son o no son modos de vida buena, cuál(es) de ellas es más importante y, por tanto, qué criterios tienen prioridad deliberativa.

Asumir esta pluralidad del valor es un gesto filosófico importante a la hora de resolver la pregunta socrática, un gesto que el paradigma moral kantiano simplemente no llega a realizar. El punto de vista de la moralidad kantiana desatiende al hecho de que el universo ético es más amplio que el universo moral. La moralidad es solamente una forma específica de vida ética, pero no es necesariamente la única. Es por esta razón que, desde el paradigma ético, resulta tan inquietante la idea presente en el sistema kantiano de que aquello que es importante y valioso para la vida humana, en su sentido más profundo, sea lo mismo para todos (*cf.* Wolf, 1997, p. 312).

Es justamente este el contexto en el que la idea de necesidad práctica adquiere tanto peso filosófico. Si el valor moral, desde una perspectiva normativa ética, no puede asumirse como el único ni como el valor más importante, resulta que, en términos éticos, la obligación de la necesidad práctica no-moral puede ser un contrapeso para la obligación moral. Dicho de otra manera, la pregunta por lo importante y por aquello que tiene prioridad deliberativa requiere criterios normativos y deliberativos que no necesariamente son morales. Con esto concluyo entonces que la necesidad práctica puede ser un contrapeso normativo a la moralidad y, por tanto, la tesis de la primacía normativa de la moralidad es falsa.

Conclusiones

Para esta investigación me propuse el objetivo de articular una crítica al rigorismo moral kantiano. Usualmente se suele acusar a la filosofía kantiana de ser rigorista debido a la posición radical que Kant asume con respecto a la mentira: debemos ser veraces incluso con el asesino en la puerta. Por ello inicié esta investigación examinado la postura de Kant sobre la mentira (capítulo 1). Este examen se complementó con el segundo capítulo en el que expliqué en qué sentido la filosofía moral kantiana está comprometida con el rigorismo moral (capítulo 2). Además, argumenté que el rigorismo moral está fundamentado en el modelo de la agencia del yo dividido (capítulo 3). En el cuarto capítulo, articulé la crítica al rigorismo kantiano: siguiendo a Williams, mostré que el rigorismo moral representa una amenaza para la integridad personal (Capítulo 4). Esta amenaza consiste en que las exigencias incondicionales de la moralidad pueden obligar a las personas a actuar en contra de su carácter y de sus proyectos fundamentales. La amenaza a la integridad es un problema en la medida en que puede conducir a la muerte práctica: a dejar de ser uno mismo.

Christine Korsgaard responde a esta crítica (Capítulo 5). Ella propone una conciliación entre la integridad y la filosofía moral kantiana. Su defensa supone, por un lado, abandonar el modelo del yo dividido y ofrecer, a cambio, su modelo constitucional. La defensa también supone interpretar la noción de integridad en términos de unidad. La tesis fuerte de Korsgaard es que los principios morales son constitutivos de la agencia y son ellos los que integran −unifican− las distintas identidades prácticas de las personas. En una palabra, sin una identidad moral constitutiva no sería posible la agencia. La integridad solo es posible en virtud de un compromiso con principios morales, y, por tanto, la moralidad kantiana no representa una amenaza a la integridad personal.

Finalmente, realizo una crítica a la respuesta de Korsgaard en tres momentos. Para empezar (§6.1), siguiendo a Josep Corbí, explico que el modelo constitucional de Korsgaard no escapa a la concepción del yo dividido, con lo cual cuestiono su pretensión de entender la integridad en términos de unidad. En segundo lugar (§6.2), incluso si el modelo constitucional es correcto y los principios morales son constitutivos de la agencia y por eso son irrenunciables, resulta que no son incondicionales: el compromiso con la moralidad está condicionado a que el agente tenga una razón para vivir. Esto significa que la moralidad no está necesariamente por encima de aquello de

dota de sentido nuestras vidas. De esta manera, la moralidad no tiene garantizada su prioridad normativa. En tercer lugar (§6.3) , aun si la anterior crítica es correcta, todavía es posible que el kantiano sostenga la tesis de la prevalencia, esto es, que la moralidad ha de tener prioridad normativa por encima de la mayoría de criterios y consideraciones morales (exceptuando, claro, los casos en los que la moralidad amenace las razones que tiene una persona para vivir). En contra de la tesis de la prevalencia, y sin pretensión de negar la importancia de la moralidad en la vida práctica de los seres humanos, muestro que la moralidad kantiana excluye otras fuentes de normatividad. Específicamente, muestro que esta manera de entender la moralidad desatiende los conceptos normativos de aquello que consideramos importante, y además muestro en qué sentido ésta confunde la normatividad moral con la normatividad ética. Esta confusión lleva a la filosofía práctica kantiana a negar que existen otro tipo de necesidades prácticas que también son incondicionales, aunque no sean necesariamente morales. Los proyectos fundamentales y las imposibilidades del carácter son ejemplos de necesidades prácticas que pueden implicar que alguien tenga que actuar de acuerdo con ellos, aun en contra de la moralidad. No es evidente, por tanto, que la moralidad tenga prioridad deliberativa frente a ellos.

Telón de fondo: el asesino en la puerta

Kant sostiene que debemos ser veraces con el asesino que llega a preguntarnos por el paradero del amigo que se oculta en nuestra casa. Esta respuesta se ha interpretado usualmente como la prueba del rigorismo moral kantiano. No obstante, esta afirmación merece ser matizada. Tal y como lo examiné, las fuentes directas de Kant no son concluyentes al respecto (capítulo 1). Sin embargo, un análisis del modelo de deliberación kantiano –el imperativo categórico y la derivación de deberes perfectos/imperfectos– muestra que los deberes perfectos son incondicionales. Y puesto que el deber de la veracidad es un deber perfecto, resulta que es un deber moral ser veraz con el asesino (capítulo 2).

Parte de mi interés por criticar el rigorismo moral kantiano se explica por mi interés en rechazar la conclusión a la que llega Kant en el caso del asesino en la puerta. Y creo que a partir de los resultados de la investigación he podido ofrecer una crítica a su posición. En una palabra, mi respuesta a Kant es que el deber de decirle la verdad al asesino atenta directamente contra la

integridad de de quien lo enfrenta. En efecto, la integridad personal se comprende a través del carácter de una persona (Capítulo 4). Ahora bien, si dijera la verdad, la persona que enfrenta al asesino estaría actuando en contra de los límites de su carácter personal (§6.3). Para esta persona, un límite de su carácter, es decir, aquello que no puede hacer —o incluso aquello que le es literalmente impensable hacer— es, por ejemplo, la traición a un amigo, favorecer un crimen o poner en riesgo la vida (propia o ajena). Y el punto es que ir en contra de lo que alguien considera que es imposible hacer significa renunciar a su carácter, significa traicionarse a sí mismo y a su identidad y, por tanto, este alguien estaría actuando en contra su propia integridad.

No siempre nos adecuamos a los dictados del deber moral, muchas veces experimentamos una disonancia con respecto a la moralidad; de hecho, desde la perspectiva kantiana, esta disonancia es casi connatural a la experiencia de la normatividad moral. El rigorista, que es un caso extremo, diría que debemos siempre actuar por deber, independientemente de esta disonancia −de hecho, para él, hay valor moral en actuar en contra de las inclinaciones−. Sin embargo, lo que he intentado explorar a lo largo de esta investigación es la posibilidad de que sea legítimo actuar en contra del deber moral. Por supuesto, esta posibilidad no puede ser la norma, pero tampoco puede ser verdad que la moralidad tenga *a priori* prioridad normativa por encima de valores no morales.

Me interesa aclarar entonces que criticar el rigorismo moral no significa, de ninguna manera, caer en un tipo de rigorismo ético; no se trata de trazar un falso dilema entre asumir la prioridad normativa de la moralidad o asumir la prioridad deliberativa de criterios éticos no morales. No intento reemplazar un rigorismo por otro[100]. Mi punto es más modesto: no es verdad que la moralidad deba imponerse siempre por encima de los valores o consideraciones no morales. Un ejemplo de esta conclusión, entonces, es el caso del asesino en la puerta: no es verdad que lo único que tengamos que hacer en un caso como este sea cumplir las exigencias del imperativo categórico. Con esta conclusión, creo, basta para desvirtuar el rigorismo moral kantiano.

Normatividad ética

La crítica en contra del rigorismo moral kantiano fue articulada en este trabajo a través de la

[100] Agradezco esta idea a Josep Corbí.

exposición y de la justificación de otras fuentes de normatividad que no son necesariamente morales y que cuestionan la prioridad normativa de la moralidad. Esta normatividad no moral bien podría denominarse *normatividad ética*, en la medida en que los fenómenos que examiné no son reducibles a la obligación moral.

En esta investigación me serví de la distinción entre ética y moral para señalar que el universo de la moral kantiana no agota todos los criterios normativos relevantes en la vida humana. Específicamente, consideré y me enfoqué en el fenómeno de la integridad personal (capítulo 4). ¿Cómo se justifica la normatividad práctica de este fenómeno al margen de la normatividad moral? Básicamente, como intenté mostrar, la integridad personal toma prestada la fuerza normativa de la pregunta socrática, es decir, su normatividad ética está fundamentada en la necesidad de entender cómo se ha de vivir una buena vida. La pregunta socrática es, sin duda, de naturaleza normativa, pues las respuestas dadas a ella suponen determinar qué tipo de vida es mejor que otra, y por lo tanto, exige comprometerse con un criterio normativo para juzgar qué es una buena vida y qué no lo es. La preocupación por la buena vida consituye entonces una fuente de razones normativas*:* ¿qué significaría para mi vida, por ejemplo, decirle la verdad al asesino? ¿Me alejaría esta acción de aquello que considerio una buena vida o podría llegar a corresponderse con ello? Estas preguntas son preocupaciones éticas y definitivamente tienen peso normativo, pues, en sí mismas son una norma para determinar si he de actuar/vivir de esta u otra manera. Y lo que es más interesante, insisto, es que este tipo de normatividad no es moral, en el sentido kantiano del término.

Digo todo esto porque es este el modo como se hace explícita la relación normativa entre la pregunta socrática, por un lado, y la integridad personal, por otro. Quien esté preocupado por cómo ha de vivir, ha de estar preocupado también por su integridad personal. La integridad es una noción que acoge múltiples elementos; es un centro alrededor del cual gravitan y se relacionan fenómenos como el carácter, las necesidades/imposibilidades prácticas, los proyectos fundamentales y el sentido de la vida de las personas. Es por esto que existe una sólida relación entre la pregunta socrática y la integridad personal: cuando una persona ve amenazada su integridad, porque se le exige trasgredir un límite de su carácter o traicionar un proyecto fundamental, en el fondo se ve sometida al peligro de que tal acción comprometa negativamente

su concepción acerca de lo que es una buena vida; una buena vida para él. La integridad personal, entonces, sirve de contrapeso normativo a la moralidad kantiana porque los fenómenos que esta acoge obedecen a una normatividad ética irreductible e inexpresable en términos de la normatividad moral.

Nuevos campos de investigación

Para terminar, quisiera señalar dos campos de investigación que, en mi opinión, se pueden desprender de los resultados de este trabajo y que pueden ser escenarios muy fructíferos para pensar más a fondo los problemas planteados aquí.

El modelo de la agencia: el modelo kantiano de la agencia, es decir, la concepción del yo dividido, resulta incompatible con la integridad personal (capítulo 3). Korsgaard, consciente de esta carencia, tiene el proyecto de conciliar la filosofía moral kantiana con el fenómeno de la integridad personal, y en su libro *Self-Constitution* (2009) propone el modelo constitucional de la agencia que, según ella, es apropiado para el éxito de su proyecto. Sin embargo, a su modelo subyace todavía una concepción escindida del yo (§6.2).

Me he comprometido con un rechazo a la concepción de un yo dividido, pues el concepto mismo de integridad está asociado a una unidad o totalidad, lo cual no se concilia muy bien con la idea de un agente dividido. Más aún, con la caracterización que hice del concepto de integridad también me comprometí con una imagen más compleja de los seres humanos: tenemos un carácter rico y heterogéneo y que cambia a través del tiempo; somos reflexivos, capaces de pensar sobre nosotros mismos y sobre el sentido de nuestra vida. Y mi sugerencia es que el modelo del yo dividido no da cuenta de esta imagen compleja de los seres humanos. De hecho, queda abierta la pregunta de si, en general, tiene sentido proponer un *modelo* que responda a esta riqueza de la vida humana.[101]

El conflicto de valores: el segundo campo de investigación tiene que ver con el conflicto práctico entre valores. Si el rigorismo moral es una posición incorrecta, si la prioridad normativa de la moralidad

[101] Agradezco esta precisión a Ángela Uribe.

falla, entonces es razonable esperar que exista un conflicto de valores.

El rigorismo moral kantiano tiene la ventaja –o la desventaja– de que disuelve cualquier conflicto práctico. Y aquí vale la pena trazar la distinción entre un conflicto práctico y un dilema moral. Sin pretender ser exhaustivo, la distinción puede radicar en el tipo de valores que entran en tensión. En un dilema moral, los valores en conflicto son siempre morales, de ahí que cualquier elección que tome alguien representa una falta moral. En cambio, ante el conflicto práctico una persona puede enfrentar valores morales y valores no-morales; y por tanto, no es necesario que ella falle moralmente cuando se decida por una de las opciones del conflicto.

El rigorismo moral, por su parte, rechazaría la existencia misma de conflictos prácticos, porque ningún valor no-moral puede igualarse y mucho menos imponerse por encima de la moralidad. Por otro lado, el rigorismo tampoco admitiría la posibilidad de los dilemas morales, pues, incluso entre los deberes morales existe una jerarquía –los deberes perfectos, por ejemplo, están por encima de los deberes imperfectos–, y además existe una coherencia interna entre los deberes: no puede ser que dos necesidades objetivas de la razón entren en conflicto, porque dos cosas necesarias no pueden ser contradictorias (*MC* 224).

Pero una vez rechazado el rigorismo moral kantiano y desvirtuada la primacía normativa de la moralidad (Capítulo 4 y 6), aparece en escena la posibilidad de conflictos prácticos. Es posible entonces que valores de distinta categoría –moral *vs* no-moral– representen un genuino conflicto para el agente. Este sería entonces el segundo campo de investigación que se desprende de la crítica al rigorismo moral: la existencia de los conflictos prácticos.

Una de las ideas centrales de esta investigación ha sido que la normatividad moral no es la única normatividad práctica. También he defendido, al menos parcialmente, que los valores éticos no pueden ser reducidos a un único tipo de valor moral, pues ello significaría un compromiso con el reduccionismo deontológico. En todo caso, a donde creo que podría dirigirse una investigación sobre esto último es al rechazo de un monismo valorativo. No existe un valor último –moral– al cual el resto de valores puedan ser reducidos. Por el contrario, creo que al introducir la distinción entre ética y moral, y al proponer distintas normatividades para cada uno de estos campos, resulta

más promisorio pensar en términos de pluralidad del valor.

He dicho además que la normatividad ética resulta irreductible a la normatividad de la moralidad kantiana. Esta conclusión puede ser llevada más lejos, en la medida en que es posible que el conflicto entre valores morales y no-morales llegue a ser inconmensurable. Tal es, por ejemplo, la posición que sostiene Thomas Nagel y que defiende a partir del hecho de que existen diversas fuentes de valor irreductibles (*cf.* Nagel, 1979b, p. 131). En este trabajo no he llegado hasta comprometerme con la tesis de la inconmensurabilidad del valor, pues para hacerlo habría que demostrar mucho más de lo que aquí he mostrado; por ejemplo, habría que dar cuenta de todos los valores humanos y luego explicar por qué ellos son irreductibles o incomparables entre sí. Además, la tesis de la inconmensurabilidad del valor tiene el reto de explicar qué sucede con la decisión en un conflicto práctico: si existen distinas fuentes de valor, y si ellos son inconmensurables entre sí, cómo es posible entonces decidir en caso de conflicto. Se presenta una "disparidad entre la fragmentación del valor y la singularidad [*singleness*] de la decisión" (*ibid.* p. 128).

Mi compromiso es, más bien, con la tesis de la irreductibilidad de la ética y la moralidad. Como he intentado mostrar, la necesidad del deber se opone a la necesidad práctica impuesta por la integridad y no es posible reducir la integridad al universo de la obligación moral. Esta situación solo constata el hecho de que, en el caso específico de conflicto práctico, no es posible cumplir con los dos tipos de valores; habrá que renunciar a alguno. Y lo que he tratado de sugerir es que no hay una forma predefinida en la que deba actuar la persona que enfrenta el conflicto. Con la crítica al rigorismo kantiano apunto en parte a una idea más amplia y que comparto con Nagel: no existe una guía normativa para resolver conflictos prácticos. Aunque a diferencia de lo que propone este autor, creo que contamos con criterios normativos, como *lo importante*, *las necesidades prácticas o las imposibilidades de carácter,* que ayudan a enfrentar los conflictos prácticos. O eso, al menos, es a donde creo que puede dirigirse una investigación sobre este tema.

Mientras sigamos pensando en un agente kantiano que puede hacerse cargo de cualquier conflicto práctico, no podremos atender al hecho de que la pluralidad del valor −y los posibles conflictos que puedan surgir entre ellos− son una muestra de lo compleja que es la vida humana. Es este tipo de complejidad la que el rigorista moral kantiano parece no poder entender.

Bibliografía

1. Obras de Kant

Kant, I. (1986). *Teoría y Práctica* (R. Rodríguez Aramayo, Ed.). Tecnos.

Kant, I. (1989). *Metafísica de las costumbres* (J. Conill Sancho & A. Cortina Orts, Eds.). Tecnos.

Kant, I. (1991). *Antropología en sentido pragmático*. Alianza.

Kant, I. (1997). *Lectures on Ethics* (P. Heath & J. B. Schneewind, Eds.). Cambridge University Press.

Kant, I. (2002a). *Fundamentación para una metafísica de las costumbres* (R. Rodríguez Aramayo, Ed.). Alianza.

Kant, I. (2002b). *Groundwork for the Metaphysics of Morals* (A. Wood, Ed.).

Kant, I. (2005). *Correspondencia* (M. Torrevejano & J. J. Carreras Ares, Trads.). Institución Fernando el Católico.

Kant, I. (2007). *Crítica de la razón pura* (M. Caimi, Ed.). Ediciones Colihue.

Kant, I. (2012). *Lecciones de ética* (R. Rodrígez Aramayo & C. Roldán Panadero, Eds.). Austral.

Kant, I. (2013). *Crítica de la Razón Práctica* (R. Rodríguez Aramayo, Ed.). Alianza.

2. Bibliografía y referencias

Allison, H. (1990). *Kant's theory of freedom*. Cambridge University Press.

Allison, H. (2012). *Essays on Kant*. Oxford University Press.

Altshuler, R. (2013). Practical Necessity and the Constitution of Character. En *The Moral Philosophy of Bernard Williams* (pp. 40–53). Cambridge Scholar Publishing.

Aristóteles. (1970). *Ética a Nicómaco* (M. Araujo & J. Marías, Eds.). Instituto de Estudios Politicos.

Arpaly, N. (2003). *Unprincipled Virtue*. Oxford University Press.

Arpaly, N., & Schroeder, T. (1999). Praise, blame and the whole self. *Philosophical Studies*. https://doi.org/10.1023/A:1004222928272

Bagnoli, C. (2009). Practical necessity: The subjective experience. En B. Centi & W. Huemer (Eds.), *Value and ontology* (pp. 22–43). Ontos-Verlag.

Bagnoli, C. (2017). *Constructivism in Metaethics*. The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/constructivism-metaethics/

Bauer, K., Varga, S., & Mieth, C. (Eds.). (2017). *Dimensions of Practical Necessity. "Here I Stand. I Can Do No Other"*. Springer International Publishing AG.

Baumann, P., & Betzler, M. (2004). *Practical Conflicts*. Cambridge University Press.

Beier, K. (2017). Three Ways to Understand Practical Necessity and Akrasia: Aristotle, Davidson, and Frankfurt. En *Dimensions of Practical Necessity* (pp. 135–160). Cham Springer International Publishing.

Benbaji, Y. (2001). The Moral, the Personal, and the Importance of What We Care about. *Philosophy*, *76*(297), 415–433.

Benton, R. J. (1982). Political Expediency and Lying: Kant vs Benjamin Constant. *Journal of the History of Ideas*, *43*(1), 135–144. https://doi.org/10.2307/2709165

Berlin, I. (1958). *The Crooked Timber of Humanity*. Princeton University Press.

Caimi, M., Beade, I., González, J., Marey, M., Modelo, F., Paolucci, M., Pringe, H., & Thisted, M. (2016). *Diccionario de la Filosofía kantiana*. Colihue.

Chandler, T. (2004). *Identification and Autonomy: A Meditation on the Philosophy of Harry Frankfurt*. University of Maryland.

Cholbi, M. (2013). The Constitutive Approach to Kantian Rigorism. *Ethical Theory and Moral Practice*, *16*(3), 439–448.

Cholbi, Michael. (2009). The Murderer at the Door: What Kant Should Have Said. *Philosophy and Phenomenological Research*, *79*(1), 17–46.

Corbí, J. (2012). *Morality, Self-knowledge and Human Suffering: An essay on The Loss of Confidence in the World*. Routledge.

Corbí, J. (2015). Principios, atención y carácter: Una defensa del particularismo moral. En P. Luque (Ed.), *Particularismo. Ensayos de filosofía del derecho y filosofía moral*. Marcial Pons.

Cox, D., La Caze, M., & Levine, M. (2017). Integrity. *The Stanford Encyclopedia of Philosophy*.

Dietrichson, P. (1964). When Is a Maxim Fully Universalizable? *Kant- Studien*, *55*, 143–70.

Dimensions of Practical. (s. f.).

Engstrom, S. (1988). Conditioned Autonomy. *Philosophy and Phenomenological Research*, *48*(3), 435–453. https://doi.org/10.2307/2107472

Enoch, D. (2006). Agency, shmagency: Why normativity won't come from what is constitutive of agency. *Philosophical Review*, *115*, 31–60.

Foot, P. (1972). Morality as a System of Hypothetical Imperatives. *The Philosophical Review*, *81*, 305–316.

Frankfurt, H. (1988). Freedom of the Will and the Concept of a Person. En *The Importance of What We Care About* (pp. 11–25). Cambridge University Press.

Frankfurt, H. (2006a). La importancia de lo que nos preocupa. Ensayos filosóficos. En *La importancia de lo que nos preocupa* (pp. 119–138). Katz.

Frankfurt, H. (2006b). Necesidad y deseo. En *La importancia de lo que nos preocupa* (pp. 153–169). Katz.

Frankfurt, H. (2007). Autonomía, necesidad y amor. En *Necesidad, volición y amor* (pp. 205–222). Katz.

Guignon, C. (2004). *On Being Authentic*. Routledge.

Hallich, O. (2017). "I Cannot Forgive You." The Unforgivable as an Example of a Practical Necessity. En *Dimensions of Practical Necessity* (pp. 37–58). Springer International Publishing.

Hallie, P. (1979). *Lest Innocent Blood Be Shed: The Story of the Village of Le Chambon and How Goodness Happened There*. Harper Perennial.

Heidegger, M. (2012). *Ser y tiempo* (J. E. Rivera Cruchaga, Trad.; Tercera edición). Editorial Trotta.

Herman, B. (1983). Integrity and Impartiality. *The Monist*, *66*(2), 233–250.

Herman, B. (1993a). Moral Deliberation and the Derivation of Duties. En *The Practice of Moral Judgment* (pp. 132–158). Harvard University Press.

Herman, B. (1993b). Obligation and Performance. En *The Practice of Moral Judgment* (pp. 159–183). Harvard University Press.

Herman, B. (1993c). *The Practice of Moral Judgment*. Harvard University Press.

Hill, T. (1971). Kant on Imperfect Duty and Supererogation. *Kant- Studien*, *62*, 55–76.

Hill, T. (1991). Autonomy and Benevolet Lies. En *Autonomy and Self-Respect*. Cambridge University Press.

Hill, T. (1992). Making Exceptions without Abandoning the Principle: Or How a Kantian Should Think about Terrorism. En *Dignity and Practical Reason in Kant's Moral Theory* (pp. 196–225). Cornell University Press.

Hobbes, T. (1996). *Leviatán* (C. Mellizo, Ed.). Alianza.

Hochberg, G. M. (1974). The Concept of "Possible Worlds" and Kant's Distinction between Perfect and Imperfect Duties. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *26*(3/4), 255–262.

Hoyos, L. E. (2014). *Ensayos de filosofía práctica y de la acción*. Universidad Nacional de Colombia - Unibiblos.

Hume, D. (1976). *Tratado de la naturaleza humana* (F. Trad. Duque, Ed.). Editora Nacional.

Isenberg, A. (1964). Deontology and the Ethics of Lying. *Philosophy and Phenomenological Research*, *24*(4), 463–480.

Jenkins, M. (2006a). *Bernard Williams*. Acumen.

Jenkins, M. (2006b). Critique of the morality system. En *Bernard Williams* (pp. 53– 86). Acumen.

Kemp, J. (1958). Kant's Examples of the Categorical Imperative. *The Philosophical Quarterly*, *8*(30).

Korsgaard, C. (1985). Kant's Formula of Universal Law. *Pacific Philosophical Quarterly*, *66*, 24–47.

Korsgaard, C. (1996). *The Sources of Normativity*. Cambridge University Press.

Korsgaard, C. (1998). Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind. *Ethics*, *109*(1), 49–66.

Korsgaard, C. (1999). Self-Constitution in the Ethics of Plato and Kant. *The Journal of Ethics*, *3*(1), 1–29.

Korsgaard, C. (2000a). *Las fuentes de la normatividad*. Universidad Autónoma de México.

Korsgaard, C. (2000b). The Right to Lie: Kant on Dealing with Evil. En *Creating the Kingdom of Ends* (pp. 133–158). Cambridge University Press.

Korsgaard, C. (2000c). Two Arguments against Lying. En *Creating the Kingdom of Ends* (pp. 335–362). Cambridge University Press.

Korsgaard, C. (2009). *Self-Constitution. Agency, Identity, and Integrity*. Oxford University Press.

Korsgaard, C. (2011). *La creación del reino de los fines*. Universidad Autónoma de México.

Locke, J. (2005). *Ensayo sobre el entendimiento humano*. Fondo de Cultura Económica.

McCarty, R. (2012). The Right to Lie: Kantian Ethics and the Inquiring Murderer. *American Philosophical Quarterly*, *49*(4), 331–343.

Mieth, C. (2017). Morality and Happiness: Two Precarious Situations? En *Dimensions of Practical Necessity* (pp. 237–252). Springer International Publishing.

Nagel, T. (1979a). Moral Luck. En *Mortal Questions*. Cambridge University Press.

Nagel, T. (1979b). The Fragmentation of Value. En *Mortal Questions* (pp. 128–141). Cambridge University Press.

Nietzsche, F. (2011). *Así habló Zaratustra*. Alianza.

Nussbaum, M. (1995). Fortuna y ética. En *La fragilidad del bien. Fortuna y ética en la tragedia y la filosofía griega* (pp. 27–50). Visor.

Oakley, J. (1990). A Critique of Kantian Arguments against Emotions as Moral Motives. *History of Philosophy Quarterly*, *7*(4), 441–459.

O'Neill, O. (1975). *Acting on principle: An essay on Kantian ethics*. https://doi.org/10.1017/CBO9781139565097

Parfit, D. (1987). *Reasons and Persons*. Oxford University Press.

Parra, L. (2017). Filosofía versus barroco en la Fundamentación de la metafísica de las costumbres de Kant. En R. R. Aramayo & F. Rivera (Eds.), *La filosofía práctica de Kant* (pp. 19–

54). Universidad Autónoma de México.

Paton, H. J. (1947). *The Categorical Imperative*. University of Pennsylvania Press.

Paton, H. J. (1954). An alleged right to lie a problem in kantian ethics. *Kant- Studien*, *45*(1–4), 190–203. https://doi.org/10.1515/kant.1954.45.1-4.190

Platón. (1988). *Diálogos IV. República*. Gredos.

Pyle, A. (1994). Interview with Bernard Williams. *Cogito*, *8*(1), 3–19.

Rawls, J. (1980). Kantian Constructivism in Moral Theory: The Dewey Lectures 1980. *Journal of Philosophy*, *77*(9), 515–572.

Rawls, J. (2006). *Teoría de la justicia*. Fondo de Cultura Económica.

Rigorism. (1910). En *New Catholic Dictionary*. https://www.studylight.org/dictionaries/ncd/r/rigorism.html

Rivera, F. (2014). *Virtud, felicidad y religión en la filosofía moral de Kant*. Universidad Autónoma de México.

Rivera, F. (2018). La primera fórmula del imperativo categórico. En *La filosofía práctica de Kant* (pp. 55–82). Universidad Autónoma de México; Instituto de Investigaciones filosóficas.

Rivera, L. (2007). Sacrifices, Aspirations and Morality: Williams Reconsidered. *Ethical Theory and Moral Practice*, *10*(1), 69–87.

Rodrígez Aramayo, Roberto. (2012). Introducción: La cara oculta del formalismo ético. En *Lecciones de ética*. Austral.

Rogerson, K. (1983). Williams and Kant on Integrity. *Dialogue*, *22*(3), 461–478.

Rosati, C. (2003). Agency and the Open Question Argument. *Ethics*, *113*, 490–527.

Rosell, S. (2009). *Carácter, circunstancias y acción. El papel de la suerte en la determinación de la responsabilidad moral.* [PhD Thesis]. Universitat de Valencia.

Sartre, J.-P. (2017). *El existencialismo es un humanismo*. Edhasa.

Schapiro, T. (2006). Kantian Rigorism and Mitigating Circumstances. *Ethics*, *117*(1), 32–57.

Schechtman, M. (2017). Loving Eyes of My Own: Love, Particularity, and Necessity. En

Dimensions of Practical Necessity (pp. 13–36). Springer International Publishing.

Schopenhauer, A. (2012). *El mundo como voluntad y representación* (Roberto R. Aramayo, Trad.). FCE.

Sedgwick, S. (1991). On Lying and the Role of Content in Kant's Ethics. *Kant- Studien*, *82*, 42–62.

Singer, M. (1954). The Categorical Imperative. *Philosophical Review*, *63*, 577–591.

Sófocles. (2000). *Tragedias* (A. Alamillo, Trad.). Gredos.

Stern, R. (2015). *Kantian Ethics. Value, Agency, and Obligation*. Oxford University Press.

Stocker, M. (1976). The Schizophrenia of Modern Ethical Theories. *The Journal of Philosophy*, *73*(14), 453–466.

Strawson, P. (1959). *Individuals*. Methuen.

Tarantino, Q. (2009). *Inglourious Bastards*. Universal Pictures.

Velleman, D. (2004). Precis of The Possibility of Practical Reason. *Philosophical Studies*, *121*, 225–238.

Williams, B. (1972). *Morality. An Introduction to Ethics*. Cambridge University Press.

Williams, B. (1973a). A Critique of Utilitarianism. En *Utilitarianism: For and Against* (pp. 77–150). Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.004

Williams, B. (1973b). Ethical Consistency. En *Problems of the Self* (pp. 166–186). Cambridge University Press.

Williams, B. (1981a). Internal and External Reasons. En *Moral Luck* (pp. 101–113). Cambridge University Press.

Williams, B. (1981b). *Moral Luck*. Cambridge University Press.

Williams, B. (1981c). Moral Luck. En *Moral Luck* (pp. 20–39). Cambridge University Press.

Williams, B. (1981d). Persons, character and morality. En *Moral Luck* (pp. 1–19). Cambridge University Press.

Williams, B. (1981e). Utilitarianism and moral self-indulgence. En *Moral Luck* (pp. 40–53). Cambridge University Press.

Williams, B. (1993a). La necesidad práctica. En *La fortuna moral* (pp. 159–167). Universidad Autónoma de México.

Williams, B. (1993b). Utilitarismo y autoindulgencia moral. En *La fortuna moral* (pp. 59–75). Universidad Autónoma de México.

Williams, B. (1995). Replies. En *World, Mind, and Ethics*. Cambridge University Press.

Williams, B. (2006a). *Ethics and the Limits of Philosophy*. Routledge.

Williams, B. (2006b). Morality, the Peculiar Institution. En *Ethics and the Limits of Philosophy* (pp. 174–196). Routledge.

Williams, B. (2006c). Socrates' Question. En *Ethics and the Limits of Philosophy* (pp. 1–21). Routledge.

Williams, B. (2006d). *Verdad y Veracidad*. Tusquets.

Williams, B. (2016a). *La ética y los límites de la filosofía* (S. Rosell, Trad.). Cátedra Teorema.

Williams, B. (2016b). La peculiar institución de la moral. En S. Rosell (Trad.), *La ética y los límites de la filosofía* (pp. 215–240). Cátedra Teorema.

Wolf, S. (1997). Meaning and Morality. *Proceedings of the Aristotelian Society*, *97*, 299–315.

Wolf, S. (2012). 'One Thought Too Many': Love, Morality, and the Ordering of Commitment. En U. Heuer & G. Lang (Eds.), *Luck, Value, and Commitment* (pp. 71–92). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199599325.003.0004

Wood, A. (1972). Kant on False Promises. En *Proceedings of the Third International Kant Conference* (Vol. 4, pp. 614–619). D. Reidel Publishing Company.

Wood, A. (2008a). *Kantian Ethics*. Cambridge University Press.

Wood, A. (2008b). Lies. En *Kantian Ethics* (pp. 240–258). Cambridge University Press.