Selecciones

Prudencia, moralidad y el Dilema del Prisionero

Derek Parfit

Hay muchas teorías sobre las razones que tenemos para actuar. Algunas de esas teorías, en ciertos casos, se destruyen a sí mismas directamente. ¿Qué muestra esto?

I

Consideremos en primer lugar el Dilema del Prisionero. Tú y yo somos interrogados por separado sobre un crimen cometido en común. Los resultados serían estos:

TU

YO	confesar	callar
confesar	Cada uno 10 años de cárcel.	Yo salgo libre. Tú 12 años de cárcel.
callar	Yo 12 años de cárcel. Tú sales libre.	Cada uno 2 años de cárcel.

A cada uno en particular le irá mejor confesando, haga el otro lo que haga. Pero si los dos confiesan saldrán los dos peor librados que si los dos guardan silencio.

Si cada uno hace lo que es mejor para sí, esto será peor para cada uno que si ninguno de los dos lo hace. Uno de los casos sucede cuando:

> *Condición positiva:* cada uno podría o (1) procurarse a sí mismo un beneficio, o (2) procurar al otro un beneficio mayor,

y,

> *Condición negativa:* ninguna de las dos decisiones sería de otra forma mejor ni peor para el otro.

Cuando se cumple la Condición positiva, los resultados serían estos:

		TU (1)	TU (2)
YO	(1)	Cada uno obtiene el beneficio menor.	Yo consigo ambos beneficios. TU ninguno.
	(2)	YO no obtengo ningún beneficio. TU ambos.	Cada uno obtiene el beneficio mayor.

Si añadimos la Condición negativa, el diagrama resulta:

		TU (1)	TU (2)
YO	(1)	Lo TERCERO mejor para AMBOS.	Lo MEJOR para MI. Lo PEOR para TI.
	(2)	Lo PEOR para MI. Lo MEJOR para TI.	Lo SEGUNDO mejor para AMBOS.

Parte de la Condición negativa no se puede mostrar en este diagrama. No debe haber reciprocidad: debe ser verdad que la elección de ninguno de

los dos induciría al otro a tomar la misma decisión. Entonces será mejor para cada uno hacer (1) en vez de (2). Esto es así haga lo que haga el otro. Pero si los dos hacen (1) esto será peor para cada uno que si ambos hacen (2).

¿Cuándo podría no haber reciprocidad? Solamente cuando cada uno debe tomar la decisión final antes de saber lo que eligió el otro. Esto no es común. Ni garantizaría la Condición Negativa. Podría haber, por ejemplo, reciprocidad diferida. La decisión de cada uno podría afectar el ser beneficiado más tarde por el otro. Por tanto rara vez podemos saber que nos enfrentamos a un Dilema del Prisionero bipersonal.

Podemos saber con frecuencia que nos enfrentamos a una versión multipersonal. Una de ellas podría llamarse el Dilema del Samaritano. Cada uno de nosotros podría ayudar en alguna ocasión a un extranjero con un coste menor para sí mismo. Cada uno podría ser ayudado a su vez de modo similar y casi con la misma frecuencia. En comunidades pequeñas, el coste de la ayuda podría asumirse indirectamente. Si yo ayudo, podré a cambio ser ayudado en el futuro. Pero en comunidades grandes esto es improbable. Para cada uno puede ser mejor no ayudar nunca. Pero sería peor para cada uno si nadie ayudase nunca. Cada uno podría ganar no ayudando nunca, pero perdería, y perdería más, no siendo nunca ayudado.

Otro caso tiene lugar cuando:

> *Condición Positiva:* cada uno de nosotros podría, con algún coste para sí mismo, procurar a los otros una suma total mayor de beneficios [1];

y

> *Condición Negativa:* no habría efectos indirectos que anulasen estos efectos directos.

La Condición positiva se cumple con frecuencia. Y si somos numerosos, lo mismo ocurre con la negativa. Lo que cada uno hace es improbable que afecte a lo que hacen los otros.

Los ejemplos más comunes son los Dilemas del Contribuyente, que implican bienes públicos: resultados que benefician incluso a aquellos que no contribuyen a producirlos. Para cada persona puede ser cierto que, si contribuye, incrementará la suma de beneficios. Pero su parte en los beneficios del incremento podría ser muy pequeña. Su contribución podría no compensarle. Tal vez fuera mejor para cada uno no colaborar. Y esto independientemente de lo que hagan los demás. Pero sería peor para cada uno que fue-

[1] O de beneficios esperados (beneficios posibles multiplicados por las probabilidades de que este acto los produzca. En muchas de mis aseveraciones posteriores «beneficio» quiere decir «beneficio esperado».

sen menos los que contribuyen. Y si ninguno contribuyese, esto sería peor para cada uno que si todos lo hicieran.

Algunos bienes públicos necesitan contribuciones financieras. Este es el caso de las carreteras, la policía o la defensa nacional. Otros necesitan esfuerzos cooperativos. Cuando en las grandes empresas los salarios dependen de los beneficios, puede ser mejor para cada uno que los otros trabajen más intensamente, peor hacerlo él mismo. Lo mismo ocurre con los campesinos en granjas colectivas. Un tercer tipo de bien público es la evitación de un mal. Esto requiere con frecuencia auto-restricción. Tales casos podrían implicar:

> *Gente que va al trabajo:* Cada uno va más rápido si lleva su coche, pero, si todos van en su coche, cada uno va más despacio que si todos van en autobús.
>
> *Soldados:* Cada uno estará más a salvo si se da la vuelta y sale corriendo, pero si todos lo hacen, morirán más que si ninguno lo hace.
>
> *Pescadores:* Cuando se pesca en exceso, puede ser mejor para cada uno intentar pescar más, peor, si todos lo hacen.
>
> *Campesinos:* Si el país está superpoblado, puede ser mejor para cada uno tener más hijos, peor, si todos los tienen.

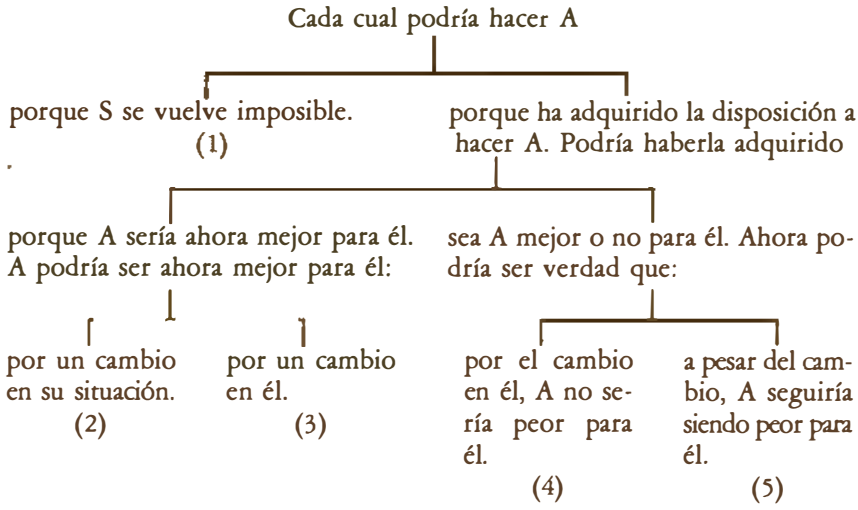
Hay muchos otros casos. Puede ser mejor para cada uno añadir contaminación, usar más energía, saltarse las colas, y romper acuerdos; pero si todos hacen esas cosas, eso puede ser peor para cada uno que si nadie lo hiciera. Con frecuencia es verdad que si cada uno, en vez de ninguno, hace lo que es mejor para él, esto será peor para todos.

II

Cada cual podría estar dispuesto a hacer lo que será mejor para él mismo. Hay pues un problema práctico. A menos que algo cambie, el resultado efectivo será peor para todos.

Usemos rótulos. Cada cual tiene dos alternativas: S (beneficiosa para sí mismo), A (altruísta). Si todos hacen S eso será para cada cual peor que si todos hacen A. Pero, hagan lo que hagan los otros, para cada cual será mejor hacer S. El problema es que, por esa razón, cada cual está ahora dispuesto a hacer S.

El problema se resolverá parcialmente si la mayoría hace A, y totalmente si todos lo hacen. Se podría alcanzar una solución de una o varias de estas formas:



Del (1) al (4) el Dilema queda abolido. La elección altruista deja de ser la peor para cada cual. Estas son con frecuencia soluciones buenas. Pero a veces son ineficaces o inalcanzables. Entonces necesitamos (5). Esto resuelve el problema práctico. Pero el Dilema no queda abolido. Subsiste un problema teórico.

En la solución (1), la elección «autobenefactora» resulta imposible. A veces esto es lo mejor. En muchos Dilemas del Contribuyente debería haber impuestos ineludibles. Pero (1) sería con frecuencia una solución pobre. Se podrían destruir las redes de pesca o encadenar a los soldados a sus puestos. Ambas soluciones tienen desventajas.

(2) es una solución menos directa. S sigue siendo posible, pero A se vuelve mejor para cada cual. Podría existir un sistema de recompensas. Pero si funciona todos deben ser recompensados. Sería mejor que la única recompensa fuera evitar una multa. Si esto funciona, nadie paga. Si se fusilase a todos los desertores no habría desertores.

(1) y (2) son soluciones políticas. Lo que cambia es nuestra situación. Del (3) al (5) son soluciones psicológicas. Somos nosotros los que cambiamos. Este cambio puede ser específico, resolviendo sólo un Dilema. Los pescadores podrían volverse perezosos, los soldados podrían preferir la muerte al deshonor. He aquí cuatro cambios de un tipo más general:

> Podríamos volvernos *fiables*. Cada cual podría entonces estar de acuerdo en hacer A a condición que los otros se asociaran a ese acuerdo.
>
> Podríamos volvernos *reacios a ser «gorrones»*. Si cada cual cree

que muchos otros harán A, podría entonces preferir cumplir su parte.

Podríamos volvernos *kantianos.* Cada cual haría sólo aquello que podría racionalmente querer que hiciesen todos. Nadie podría querer racionalmente que todos hiciesen S. Por tanto, cada cual haría A.

Podríamos volvernos *más altruistas.* Dado el suficiente altruísmo, cada cual haría A.

Estas son soluciones morales. Puesto que podrían resolver cualquier Dilema, son las soluciones psicológicas más importantes.

Son con frecuencia mejores que las soluciones políticas. Y esto en parte porque no tienen que ser impuestas. Tomemos el Dilema del Samaritano. No puede hacerse imposible no ayudar a los extraños. Los malos samaritanos no pueden ser fácilmente capturados y multados. Se podría recompensar a los buenos samaritanos. Pero para asegurar esto tendría que intervenir la ley. Dados los costes administrativos esta solución podría no valer la pena. Sería mucho mejor que adquiriésemos directamente la disposición a ayudar a los extraños.

No basta con saber qué solución sería la mejor. Toda solución ha de ser introducida. A menudo es más fácil lograrlo con las soluciones políticas. Es más fácil cambiar las situaciones que a la gente. Pero a menudo nos encontramos con otro Dilema del contribuyente. Pocas soluciones políticas pueden ser introducidas por una persona sola. La mayoría requieren la cooperación de mucha gente. Pero una solución es un bien público, que beneficia a cada uno, contribuya o no a producirlo. En la mayoría de los grupos grandes, para cada uno no será lo mejor aportar su propia parte. Su propia contribución no se notará gran cosa.

Este problema puede ser pequeño en democracias bien organizadas. Puede ser aquí suficiente hacer que el problema original sea ampliamente entendido. Esto puede ser difícil. Pero entonces podemos votar por una solución política. Con un gobierno que responda podría no haber necesidad de convocar elecciones.

El problema es mayor cuando no hay ningún gobierno. Esto es lo que preocupaba a Hobbes. Un ejemplo es la proliferación de armas nucleares. Sin un gobierno mundial, puede ser difícil encontrar una solución.

El problema es mayor todavía cuando a su solución se opone algún grupo dominante. Este es el *Dilema del Oprimido.*

Dilemas tales como el del contribuyente requieren con frecuencia soluciones morales. A menudo necesitamos de alguna gente directamente dispuesta a cumplir con su parte. Si estos pueden cambiar la situación, alcanzando una solución política, ésta puede sostenerse por sí misma. Pero sin esa gente podría no alcanzarse nunca.

Las soluciones morales son, pues, con frecuencia, las mejores; y con frecuencia son las únicas alcanzables. Por tanto, necesitamos motivos morales. ¿Cómo se pueden introducir? Afortunadamente, ese no es problema nuestro. Existen. Así es como resolvemos muchos Dilemas del Prisionero. Lo que necesitamos es reforzar esos motivos y difundirlos más ampliamente.

Para esta tarea ayuda la teoría. El Dilema del Prisionero tiene que ser explicado. Sus soluciones morales tienen que serlo también. Ambos han sido demasiado poco comprendidos.

Una solución es, como vimos, un acuerdo condicional. Para que sea posible, debe ser cierto primero que todos podamos comunicarnos. Si estamos interesados por nosotros mismos, es raro que esto importe. En la mayor parte de los grupos grandes no tiene mayor sentido acordar que haremos la elección altruista, puesto que es mejor para cada uno romper el acuerdo. Pero supongamos que somos fiables. Cada cual podría prometer hacer A, a condición de que todos los otros hagan la misma promesa. Si sabemos que todos somos fiables, cada uno tendrá un motivo para asociarse a este acuerdo condicional. Cada uno sabrá que, a menos que se asocie, el acuerdo no tendrá efecto. Una vez que todos hemos hecho esta promesa, todos haremos A.

En casos en que hay poca gente implicada, ese tipo de acuerdo conjunto condicional puede ser una buena solución. Pero en casos que implican a gran cantidad de personas son poco útiles. Lleva algún esfuerzo tanto conseguir que todos se comuniquen como alcanzar un acuerdo mutuo. Pero el acuerdo es un bien público, que beneficia a cada uno, ayude o no a producirlo. En grupos muy grandes, ayudar no resulta mejor para cada uno en particular. La fiabilidad no aporta ninguna solución a este Dilema del contribuyente.

Si somos reacios a ser gorrones este problema se reduce. No hay ahora necesidad de acuerdo alguno. Todo lo que se necesita es la seguridad de que habrá muchos que hagan A. Cada cual preferiría entonces cumplir su parte. Pero la repugnancia a ser gorrones no puede crear por sí misma esa seguridad. Hay, pues, muchos casos en los que no aporta ninguna solución.

El «test» kantiano siempre podría proporcionar una solución. Este «test» tiene sus propios problemas. ¿Podría yo racionalmente querer, bien que nadie practicase la medicina, bien que todos lo hiciesen? Si refinamos el «test» podríamos resolver dichos problemas. Pero estos no se plantean en el Dilema del Prisionero. Esos son los casos en los que decimos naturalmente «¿Qué pasaría si todos lo hiciesen?».

La cuarta solución es suficiente altruismo. Esta ha sido la menos comprendida. Cada elección altruista beneficia a los otros. Pero en los Dilemas del contribuyente el beneficio para cada uno de los otros puede ser muy pequeño. Incluso puede ser imperceptible. Algunos creen que tales beneficios no tienen ninguna relevancia moral. Si esto fuera así, los altruístas racionales no contribuirían.

No puede ser así. Consideremos la Paradoja del donante. Muchos hombres heridos yacen en el desierto. Cada uno de nosotros tiene un cuartillo de agua, que podría transportar a alguno de los heridos. Pero si nuestros cuartillos se transportan separadamente, se evaporará gran parte del agua. Si en cambio echamos los cuartillos en un camión cisterna no habrá evaporación. Para altruístas racionales éste sería el mejor modo de dar. Cada herido recibiría más agua. Pero el cuartillo que cada uno de nosotros aporta sería ahora repartido entre todos esos muchos hombres. Se daría a cada uno de ellos solamente una única gota. Incluso para un hombre herido cada gota de agua es un beneficio muy pequeño. Si ignoramos tales beneficios, deberemos concluir que cada una de nuestras contribuciones ha sido malgastada.

Subdividimos a continuación las soluciones morales. Cuando algún motivo moral lleva a alguien a hacer A, lo que hace puede ser o no ser peor para él. Esta distinción plantea una profunda cuestión. Pero yo expondré simplemente lo que mi argumento supone. Lo que nos interesa depende parcialmente de cuáles sean nuestros motivos. Si tenemos motivos morales, puede no ser cierto que hacer A sea peor para nosotros. Pero podría serlo. Incluso sabiéndolo podríamos, sin embargo, hacer A.

Estoy descartando aquí cuatro afirmaciones. Algunos dicen que nadie hace lo que cree que es peor para él. Esto se ha refutado con frecuencia. Otros dicen que lo que cada uno hace es, por definición, lo mejor para él. En frase de los economistas, «maximizará su utilidad». Como se trata de una mera definición no puede ser falso. Pero es aquí irrelevante. Simplemente no trata de los intereses a largo plazo de la persona. Otros dicen que la virtud es siempre recompensada. A menos que haya otra vida, esto ha sido refutado también. Otros dicen que la virtud es su propia recompensa. Esto es demasiado oscuro para ser rechazado fácilmente —o para ser discutido aquí.

Volviendo a mi propia exposición. Muchos Dilemas del Prisionero requieren soluciones morales. Debemos llegar a estar directamente dispuestos a adoptar decisiones altruístas. Esas soluciones son de dos tipos. Algunas eliminan el Dilema. En tales casos, a causa del cambio producido en nosotros, deja de ser cierto que sea peor para cada uno hacer A. Pero en otros casos sigue siéndolo. Incluso en tales casos podríamos hacer A. Cada cual podría hacer, por razones morales, lo que sabe que es peor para él.

Con frecuencia necesitamos soluciones morales de este segundo tipo. Llamémoslas «ab-negadas». Estas solucionan el problema práctico. El resultado es mejor para todos. Pero no eliminan el Dilema. Queda en pie un problema teórico.

III

Es éste. Podemos tener razones morales para hacer A. Pero resultará mejor para cada cual el hacer S. La moralidad entra en conflicto con el propio interés. Cuando aparece este conflicto, ¿qué es racional hacer?

Desde un cierto punto de vista lo racional es la elección en favor del beneficio propio. Este punto de vista carece de nombre adecuado. Llamémoslo prudencia. Si aceptamos este punto de vista, seremos ambivalentes acerca de las soluciones morales «autonegadoras». Creeremos que, para alcanzar tales soluciones, todos debemos actuar irracionalmente.

Muchos autores se resisten ante esta solución. Algunos afirman que las razones morales no son más débiles que las razones prudenciales. Otros afirman, más atrevidamente, que son más fuertes. Desde su punto de vista, la elección racional es la altruísta.

Este debate puede parecer irresoluble. ¿Cómo pueden contrapesarse esos dos tipos de razones? Las razones morales son, desde luego, moralmente supremas. Pero las razones prudenciales son prudencialmente supremas. ¿Dónde podemos encontrar una escala neutra?

Algunos creen que no necesitamos una escala neutra. Afirman que en los Dilemas del Prisionero la prudencia se autodestruye. Incluso en términos prudenciales, la moralidad vence.

¿Es así? Llamemos a la prudencia

> *individualmente autodestructiva* cuando sea peor para alguien el ser prudente,

y

> *colectivamente autodestructiva* cuando sea peor para cada uno el que todos, más bien que ninguno, sean prudentes.

La prudencia podría ser individualmente autodestructiva. Podría ser cierta cualquiera de las proposiciones siguientes:

> (1) Podría ser peor para alguien actuar prudentemente. Cuando hay incertidumbre, el acto prudente puede no ser el que resulte mejor.
>
> (2) Podría ser peor para alguien estar dispuesto a actuar prudentemente. Podría ser peor para él incluso aunque siempre hiciese lo mejor para él. Un ejemplo es la «paradoja del hedonismo»: poner como meta la felicidad puede hacerla más difícil de alcanzar.

En el Dilema del Prisionero ninguna de esas proposiciones es cierta. En él los malos efectos son producto de actos, no de disposiciones. Y no existe

incertidumbre. Será mejor para cada uno actuar prudentemente; y hagan lo que hagan los otros, lo mejor para cada cual será realizar esta opción. Resulta así que la prudencia no es en este caso individualmente autodestructiva. Pero lo es colectivamente. Que todos actúen prudentemente será para cada uno peor que si ninguno lo hace.

¿Muestra esto que, si todos actuamos prudentemente, somos irracionales? Podemos partir de una cuestión más limitada. ¿Lo muestran así nuestros propios supuestos? ¿Falla nuestra prudencia incluso en sus propios términos?

Podemos responder: «No. La prudencia de cada uno es mejor para él. Tiene éxito. ¿Por qué es nuestra prudencia colectivamente autodestructiva? Sólo porque la prudencia de cada uno es peor para los demás. Esto no la hace infructuosa. No es benevolencia». Si somos prudentes, deploraremos, desde luego, los Dilemas del Prisionero. No son éstos los casos preferidos por los economistas clásicos, en los que cada cual gana gracias a la universal prudencia. Podríamos decir: «En esos casos la prudencia funciona y además aprueba la situación. En los Dilemas del Prisionero, la prudencia sigue funcionando. Cada uno sigue ganando gracias a su propia prudencia. Pero como cada uno pierde más a causa de la prudencia de los demás, la prudencia aquí condena la situación».

Esto podría parecer una evasión. Cuando es peor para cada uno que todos seamos prudentes, podría parecer que nuestra prudencia debería condenarse a sí misma. Supongamos que en algún otro grupo enfrentado con el mismo Dilema todos realizan la elección altruísta. Podrían decirnos: «Nos creéis irracionales. Pero nos va mejor que a vosotros. Actuamos mejor incluso en términos prudenciales».

Podríamos responder: «Eso es sólo un juego de palabras. "Actuáis mejor" sólo en el sentido de que os va mejor. Cada uno de vosotros *actúa* peor en términos prudenciales. Hace lo que es peor para él». Y podríamos añadir: «Lo que es peor para cada uno de nosotros es que, en nuestro grupo, no hay tontos. Cada uno de vosotros tiene mejor suerte. Su propia irracionalidad es peor para él, pero gana más aún gracias a la irracionalidad de los otros».

Ellos podrían responder: «Tenéis en parte razón. Cada uno de nosotros *actúa* peor en términos prudenciales. Pero aunque *cada uno* actúa peor, *nosotros* actuamos mejor. No es un juego de palabras. A cada uno nos va mejor a causa de lo que *hacemos*».

Esta sugerencia parece más prometedora. Volvamos al caso más simple de dos personas. Cada uno podría, o bien beneficiarse a sí mismo (S), o bien dar al otro algún beneficio mayor (A). Los resultados serían estos:

		TU	
		haces S	haces A
YO	hago S	Lo TERCERO mejor para AMBOS.	Lo MEJOR para MI. Lo PEOR para TI.
	hago A	Lo PEOR para MI. Lo MEJOR para TI.	Lo SEGUNDO mejor para AMBOS.

Para asegurar que la elección de ninguno de los dos pueda afectar a la del otro, supongamos que no podemos comunicarnos. Si yo hago A en vez de S, resultará peor para mí. Será así hagas lo que hagas. Y lo mismo te pasa a ti. Si ambos hacemos A en vez de S, cada uno hace, en consecuencia, lo peor en términos prudenciales. La sugerencia es que nosotros actuamos mejor.

Lo que la hace prometedora es que contrasta «cada uno» con «nosotros». En algunas afirmaciones ambos son equivalentes. No puede ser verdad que *cada uno* es viejo, pero que *nosotros* somos jóvenes. Pero en otras no lo son. Puede ser cierto que *cada uno* es débil, pero que *nosotros* somos fuertes. Nosotros *juntos* podríamos serlo. Nuestra sugerencia es de este segundo tipo. Podría ser verdad que, aunque cada uno esté actuando peor en términos prudenciales, nosotros juntos estemos actuando mejor.

¿Es cierto? Usemos esta prueba. Nuestra prudencia nos fija a cada uno determinado objetivo. *Cada uno* actúa mejor, en términos prudenciales si alcanza más eficazmente este objetivo. *Nosotros* actuamos mejor, en los mismos términos, si alcanzamos más eficazmente el propósito de cada uno. Esta prueba parece justa. Podría mostrar que, si *cada uno* actúa lo mejor que puede *nosotros* juntos no podríamos actuar mejor.

¿Cuál es el objetivo que nuestra prudencia da a cada uno? Podríamos decir, «actuar prudentemente». Esto es verdad, pero engañoso. Algunos objetivos son fundamentales. Otros se derivan de éstos. Llamemos a los primeros metas. Cuando medimos el éxito, sólo cuentan las metas. Supongamos que estamos intentando rascarnos la espalda. La meta de cada cual podría ser que nos deje de picar. Actuaríamos mejor, entonces, si cada uno rascase la espalda al otro. Pero podríamos ser contorsionistas: la meta de cada uno podría ser rascarse a sí mismo la espalda. Si nos rascásemos mutuamente la espalda, entonces actuaríamos peor.

Si somos prudentes, ¿cuál es la meta de cada uno? ¿Es promover su interés, o promoverlo *él*? Si fuera lo segundo, no seríamos prudentes. Tal vez somos nietzscheanos, cuyo ideal es «la más acérrima dependencia de sí mismo». Si ambos hacemos A en vez de S actuaríamos peor, según esos términos. El interés de cada uno se promovería mejor. Pero el de ninguno de ambos sería

promovido por él mismo. Ninguno de los dos alcanzaría su meta.

Este ideal nietzscheano no es la prudencia. Ambos fijan a cada uno el objetivo de la propia promoción. Pero sólo para los nietzscheanos es ésta la meta. Para el prudente, cualquier acto es un mero medio. La meta es siempre el efecto —sea éste el placer o algún otro beneficio. (Se dice que las «bestias rubias» de Nietzsche eran leones. Pero para ellos, también, actuar es un medio. Ellos prefieren comer lo que otros matan).

La meta de la prudencia de cada persona es el mejor resultado posible para sí misma. Si ambos hacemos A en vez de S, logramos un resultado mejor para cada uno. Hacemos que la meta de cada uno se alcance mejor. Por ello actuamos mejor en términos prudenciales. Esto confirma lo que sugerimos antes. El acto prudente es S. Si ambos actuamos prudentemente actuamos peor de lo que podríamos incluso en términos prudenciales.

¿Muestra esto que la prudencia se condena a sí misma? Podría parecerlo. Y es tentador contrastar la prudencia con la moralidad. Podríamos decir: «La prudencia engendra el conflicto al decirle a cada uno que trabaje contra los demás. Así es como la prudencia universal puede ser peor para todos. Donde la prudencia divide la moralidad une. Esta nos dice que trabajemos juntos —que hagamos lo mejor que *nosotros* podamos. Por ello, incluso dentro de la escala que proporciona el propio interés la moralidad gana. Esto es lo que nos enseñan los Dilemas del Prisionero. Si trocamos prudencia por moralidad, actuamos mejor incluso en términos prudenciales».

Pero esto es demasiado precipitado. *Nosotros* actuamos mejor, pero *cada uno* actúa peor. Si ambos hacemos A en vez de S, *nosotros* logramos un resultado mejor para cada uno, pero *cada uno* logra un resultado peor para sí mismo. Haga el otro lo que haga, sería mejor para cada uno hacer S. En los Dilemas del Prisionero el problema es éste. ¿Debería *cada uno* hacer lo mejor que pueda para sí mismo? ¿O deberíamos *nosotros* hacer lo mejor que podamos para cada uno? Si *cada uno* hace lo mejor para sí, *nosotros* actuamos peor de lo que podríamos para cada uno. Pero *nosotros* hacemos lo mejor para cada uno sólo si *cada uno* actúa peor de lo que podría para sí mismo.

Este es sólo un caso especial de un problema más amplio. Consideremos una teoría sobre las razones que tenemos para actuar. Puede haber casos en los que, si *cada uno* hace lo mejor en términos de esta teoría, *nosotros* hacemos lo peor, y viceversa. Llamemos a estos casos «Dilemas del Cada Uno/Nosotros».

Algunas teorías no pueden producir tales dilemas. Luego veremos por qué ocurre así en ciertas teorías. No es evidente qué es lo que muestra el hecho de que una teoría produzca Dilemas de Cada Uno/Nosotros. Consideremos nuevamente la prudencia. Esta dice a cada uno que haga lo que mejor pueda para sí mismo. Estamos discutiendo casos en los que, si todos actuamos prudentemente, hacemos lo peor para cada uno. La prudencia es aquí colectivamente autodestructiva. Pero no es evidente que esto constituya un defecto. ¿Por qué habría de ser una teoría colectivamente venturosa? ¿Por

qué no basta que funcione en el nivel individual?

Podríamos decir: «Pero una teoría no puede aplicarse tan sólo a un único individuo. Si es racional para mí actuar prudentemente, tiene que ser racional para todos hacerlo así. Cualquier teoría aceptable ha de ser por tanto venturosa en el nivel colectivo».

Esto implica una confusión. Llamemos *universal* a una teoría si se aplica a todos, y *colectiva* si se reclama el éxito en el nivel colectivo. Algunas teorías poseen ambas características. Un ejemplo es una moralidad kantiana. Este dice a cada uno que haga sólo aquello que pueda racionalmente querer que todos hagan. Los planes o políticas de cada uno han de ser contrastados en el nivel colectivo. Para un kantiano, la esencia de la moralidad es el movimiento desde *cada uno* a *nosotros.*

En el nivel colectivo —como respuesta a la pregunta «¿cómo deberíamos actuar todos nosotros?»— la prudencia se *condenaría* a sí misma. Supongamos que estamos eligiendo el código de conducta que se fomentará públicamente o se enseñará en las escuelas. Aquí sería prudente votar contra la prudencia. Si estamos eligiendo un código colectivo, la elección prudente sería la moralidad.

La prudencia es una teoría universal aplicable a todos. Pero no es un código colectivo. Es una teoría de la racionalidad individual. Esto responde a la cuestión más limitada que planteamos antes. En el Dilema del Prisionero, en el que la prudencia es autodestructiva sólo colectivamente, ésta no se condena a sí misma.

IV

Muchas malas teorías no se condenan a sí mismas. Así, la cuestión más amplia permanece abierta. En tales casos, ¿cuál es la acción racional?

Puede ayudar el introducir otra teoría común. Esta dice a cada uno que haga aquello que logre mejor sus objetivos presentes. Llamémosla la *teoría instrumental.* Supongamos que, en algún Dilema del Prisionero, mi objetivo es el resultado mejor para mí. Según la teoría instrumental, es entones la elección prudente la que es racional. Si mi objetivo es beneficiar a otros o aplicar la prueba kantiana, la racional es la elección altruísta. Si mi objetivo es hacer lo que otros hacen —quizá porque no deseo ser gorrón— es incierto qué elección es la racional. Esto depende de mis creencias acerca de lo que los otros hacen.

Como muestran estas observaciones, la teoría instrumental puede entrar en conflicto con la prudencia. Lo que mejor consigue mi objetivo presente, puede estar en contra de mi propio interés a largo plazo. Puesto que las dos teorías pueden entrar en conflicto, aquellos que creen en la prudencia deben rechazar la teoría instrumental.

Estos podrían advertir que, incluso en el nivel individual, la teoría instrumental puede ser autodestructiva. Puede producir Dilemas intertemporales. Estos serán muy frecuentes si me preocupa menos mi futuro ulterior. Supongamos que, en tiempos diferentes, tengo objetivos conflictivos entre sí. En cada momento puedo o (1) hacer aquello que logrará mejor mis objetivos presentes, o (2) hacer aquello que logrará mejor, o me hará capaz de lograr todos mis objetivos a lo largo del tiempo. Según la teoría instrumental, yo debería hacer siempre (1), más bien que (2). Sólo así haré en cada momento lo mejor que puedo en términos instrumentales. Pero entonces, a lo largo del tiempo, podría estar actuando peor, en esos mismos términos. A lo largo del tiempo total puedo tener menos éxito en lograr mis objetivos en cada momento. (He aquí un ejemplo trivial. En cada momento lograré mejor mi objetivo presente si no malgasto energía alguna en ser ordenado. Pero si nunca soy ordenado, esto puede ser la causa de que en cada momento posterior lo logre menos).

Aquellos que creen en la prudencia pueden apelar a tales casos. Podrían decir: «La teoría instrumental es aquí autodestructiva. Incluso en los términos de esta teoría, la prudencia es superior. El acto prudente es (2). Si siempre haces (2) en vez de (1), lograrás más eficazmente tus objetivos en cada momento. Si eres prudente, actúas mejor incluso en términos instrumentales».

Esto es, una vez más, demasiado precipitado. Yo actúo mejor *a lo largo del tiempo.* Pero *en cada momento* actúo peor. Si siempre hago (2) estoy haciendo en cada momento lo que me hace lograr menos eficazmente los objetivos que tengo entonces. (1) es lo que mejor me lo hará lograr. Recordemos el Dilema interpersonal. Sustituyamos la palabra «nosotros» por «yo a lo largo del tiempo», y la palabra «cada uno» por «yo en cada momento». En el Dilema interpersonal, *nosotros* actuamos mejor sólo si *cada uno* actúa peor de lo que podría. En el Dilema intertemporal, yo actúo mejor a lo largo del tiempo solamente si en cada momento actúo peor de lo que podría entonces.

Debemos distinguir de nuevo dos niveles. La teoría instrumental es aquí *intertemporalmente* autodestructiva. Pero no pretende tener éxito en el nivel intertemporal. Por eso no se condena a sí misma. No fracasa en sus propios términos.

Los que creen en la prudencia deben afirmar que, sin embargo, debe ser rechazada. Podrían decir: «Toda teoría aceptable debe tener éxito intertemporalmente. No sirve de defensa alegar que la teoría instrumental no pretende tal éxito. Esto muestra únicamente que es estructuralmente defectuosa. Si una teoría es intertemporalmente autodestructiva, eso basta para mostrar que ha de ser rechazada».

Este argumento es peligroso. Si refuta la teoría instrumental por ser intertemporalmente autodestructiva, ¿por qué no refuta la prudencia, que lo es colectivamente? Y si es una buena réplica que la prudencia no pretende

tener éxito colectivamente, ¿por qué no puede la teoría instrumental replicar de modo similar?

Como muestra todo esto, se puede desafiar a la prudencia desde dos direcciones. Esto la hace más difícil de defender. La respuesta a uno de los desafíos puede minar las respuestas al otro.

Un desafío proviene de las teorías morales. El otro no tiene que provenir de la teoría instrumental. Puede provenir de teorías más plausibles. La teoría instrumental tiene dos características. Es *temporalmente relativa:* apela a los objetivos del agente en el momento de actuar. Y es *puramente instrumental:* discute solamente medios, tomando los objetivos del agente como dados. De acuerdo con esta teoría ningún objetivo es irracional. Cualquier objetivo puede ofrecer razones para actuar.

Otras teorías son temporalmente relativas al tiempo, pero no puramente instrumentales. Un ejemplo es la *teoría deliberativa.* Esta apela, no a los objetivos que de hecho se propone el agente en el momento de actuar, sino a los que tendría si conociese los hechos y pensase con claridad. De acuerdo con esta teoría, si un objetivo no sobreviviese a tal deliberación, no proporcionaría buenas razones para actuar. Un teórico deliberativo podría añadir ulteriores exigencias. Podría decir que, incluso si sobreviviesen a esta prueba, ciertos tipos de objetivos son intrínsicamente irracionales.

Como es temporalmente relativa, la teoría deliberativa puede estar en conflicto con la prudencia. Alguien puede estar pensando claramente y sin embargo proponerse objetivos que sabe que se oponen a su propio interés a largo plazo. Y podemos negar que con ello se muestre que todos esos objetivos son irracionales. Podemos creer que hay muchos objetivos que no son menos racionales que la prosecución del propio interés. Algunos ejemplos podrían ser: beneficiar a otros, descubrir verdades o crear belleza. Según una teoría temporalmente relativa lo que es racional que yo haga ahora, depende de cuáles sean, entre todos esos objetivos, los que poseo ahora.

Los que creen en la prudencia deben rechazar tales teorías. Deben afirmar que las razones para actuar no pueden ser temporalmente relativas. Podrían decir: «la fuerza de una razón se extiende más allá del tiempo. Puesto que *tendré* razón para promover mis objetivos futuros, tengo razón para hacerlo así *ahora*». Esta exigencia está en el corazón de la prudencia.

Muchos teóricos morales hacen una segunda exigencia. Creen que ciertas razones no son relativas al agente. Podrían decir: «La fuerza de una razón puede extenderse, no sólo a lo largo del tiempo, sino sobre diferentes vidas. De modo que si *tú* tienes una razón para aliviar tu dolor, ésa es también una razón para mí. *Yo* tengo una razón para aliviar tu *dolor*».

La prudencia acepta la primera exigencia pero rechaza la segunda. Puede ser difícil defender ambas mitades de esta posición. En réplica al moralista, el prudente puede preguntar «¿por qué debería *yo* conceder peso a objetivos que no son *míos*?». Pero entonces se le puede preguntar «¿por qué debería yo dar importancia *ahora* a fines que no son míos *ahora*?» Puede responder ape-

lando a los Dilemas intertemporales, en los que las teorías temporalmente relativas son intertemporalmente autodestructivas. Pero entonces puede retársele con los Dilemas interpersonales, en los que su propia teoría es colectivamente autodestructiva. El moralista podría decir: «el argumento en favor de la prudencia nos lleva más allá de la prudencia. Entendido propiamente es un argumento en favor de la moralidad».

Esta es una línea de pensamiento tentadora. Pero antes hay que discutir otra cosa. En el nivel interpersonal el contraste *no* es entre la prudencia y la moralidad.

V

Nos ayudarán algunas distinciones más. Hemos estado considerando diferentes teorías sobre la racionalidad. Podemos describir tales teorías indicando qué nos dicen que intentemos alcanzar. De acuerdo con todas estas teorías deberíamos intentar actuar racionalmente. Llamemos a nuestro objetivo *formal.* Podemos dejar esto de lado aquí. Por «objetivos» queremos decir objetivos *sustantivos.* Podemos describir las teorías morales del mismo modo. De acuerdo con todas esas teorías deberíamos intentar actuar moralmente. Las diferentes teorías morales nos presentan diferentes objetivos sustantivos.

A continuación podemos distinguir dos modos en los que una teoría podría ser sustantivamente autodestructiva. Llamemos a esta teoría T, y a los objetivos que nos presente *nuestros objetivos T-dados.* Digamos que *seguimos con éxito T,* cuando cada uno realiza con éxito, de entre los actos disponibles, el que mejor logra sus objetivos T-dados. Llamemos a T

> *indirectamente autodestructiva,* cuando el mejor modo de lograr nuestros objetivos T-dados sea únicamente no intentar lograrlos,

y

> *directamente autodestructiva,* cuando el mejor modo de lograr nuestros objetivos T-dados sea únicamente no tener éxito en seguir T.

Consideremos primero una teoría moral: Consecuencialismo del Acto, o CA. Esta teoría propone a todos un objetivo común: el mejor resultado posible. Si intentamos lograr este objetivo, quizá a menudo fallemos. Incluso cuando tengamos éxito, el hecho que estamos dispuestos a intentarlo podría hacer el resultado peor. CA podría ser así indirectamente autodestructiva. ¿Qué muestra esto? Un consecuencialista podría decir: «Muestra que CA debería ser solamente una parte de nuestra teoría moral. Sería la parte que

cubre los actos con éxito. Cuando poseemos la certeza del éxito, debemos proponernos como objetivo el mejor resultado posible. Nuestra teoría más amplia habría de ser ésta: debemos tener los objetivos y disposiciones cuya posesión produjese el resultado mejor. Esta teoría más amplia no sería autodestructiva. La objeción quedaría así refutada».

¿Podría ser CA *directamente* autodestructiva? ¿Podría ser cierto que produjésemos el mejor resultado sólo si no seguimos con éxito CA? No es posible. Seguimos con éxito CA cuando cada uno de nosotros realiza, de los actos disponibles, el que produce el mejor resultado. Esto no asegura que nuestros actos produzcan conjuntamente el mejor resultado posible. Pero, si lo hacen, debemos estar siguiendo con éxito CA. Así pues, CA no puede ser directamente autodestructiva.

Podemos ampliar esta conclusión. Cuando cualquier teoría T propone a todos los agentes objetivos *comunes,* no puede ser directamente autodestructiva. Si hacemos que esos objetivos comunes se logren de forma óptima, tenemos que estar siguiendo con éxito T. No puede, pues, ser verdad que logremos de forma óptima nuestros objetivos T-dados solamente si no seguimos con éxito T.

¿Qué pasa si T propone a *diferntes* agentes *diferentes* objetivos? Podría no haber forma de lograr de forma *óptima* los objetivos T-dados de *cada uno.* Así, pues, debemos cambiar nuestra definición. Y necesitamos nuestra distinción anterior. Llamemos a T

> *directa e individualmente auto-destructiva,* cuando sea cierto que, si alguien sigue con éxito T, por ello mismo hará que sus objetivos T-dados se logren peor,

y

> *directa y colectivamente auto-destructiva,* cuando sea cierto que, si todos en vez de ninguno seguimos con éxito T, por ello mismo haremos que los objetivos T-dados de cada uno se logren peor.

Supongamos que T nos propone a ti y a mí diferentes objetivos. Y supongamos que cada uno puede, o bien (1) promover su propio objetivo T-dado, o bien (2) promover más eficientemente el del otro. Los resultados serían éstos:

		TU	
		haces (1)	haces (2)
YO	hago (1)	El objetivo T-dado de cada uno es el 3.º mejor logrado.	El mío se logra mejor, el tuyo peor.
	hago (2)	El mío se logra peor, el tuyo mejor.	El objetivo T-dado de cada uno es el 2.º mejor logrado.

Supongamos finalmente que la decisión de ninguno de ambos afectará a la del otro. Para cada uno de ellos será entonces verdad que, si hace (1) en vez de (2), hará con ello que su objetivo T-dado se logre mejor. Esto es así haga lo que haga el otro. Así pues, ambos seguiremos T con éxito sólo si ambos hacemos (1) en vez de (2). Solamente entonces realiza cada uno, de los actos disponibles, el que mejor logra su objetivo T-dado. Pero es cierto que si ambos en vez de ninguno seguimos T con éxito —si ambos hacemos (1) en vez de (2)—, con ello haremos que el objetivo T-dado de cada uno se logre peor. La teoría T es aquí directa y colectivamente autodestructiva.

Si sustituimos «T» por «prudencia», acabamos de describir un Dilema del Prisionero. Como esto muestra, nada depende del contenido de la prudencia. Tales casos pueden darse cuando:

(a) la teoría T es *relativa al agente,* y propone objetivos diferentes a agentes,

(b) el logro del objetivo de cada persona depende parcialmente de lo que los otros hacen,

y

(c) lo que cada uno hace no afecta a lo que estos otros hacen.

Estas condiciones pueden cumplirse si sustituimos «moral del sentido común» por «T».

VI

Muchos de nosotros creemos que hay personas hacia las cuales tenemos obligaciones especiales. Son aquellas personas con las que estamos en deter-

minadas relaciones —tales como nuestros hijos, padres, alumnos, pacientes, miembros de nuestro mismo sindicato, o aquellos a los que representamos. Creemos que debemos ayudar a esas personas de determinadas formas. Debemos tratar de protegerlos de cierto tipo de daños y procurarles cierto tipo de beneficios. La moral del sentido común consiste en gran medida en tales obligaciones.

Cumplir con esas obligaciones tiene prioridad sobre ayudar a extraños. Esta prioridad no es absoluta. No podemos creer que debo salvar a mi hijo de algún daño menor, antes que salvar la vida de un extraño. Pero debo proteger a mi hijo antes que salvar a extraños de daños *algo* mayores. Mi obligación para con mi hijo no queda superada cada vez que yo pueda hacer algún bien mayor en otra parte.

Cuando trato de proteger a mi hijo, ¿cuál debería ser mi objetivo? ¿Debe ser simplemente que no sea dañado? ¿O más bien que sea preservado del daño por mí? Si tú tuvieras mejores probabilidades de preservarlo de algún daño, sería incorrecto por mi parte insistir en ser yo quien lo intente. Esto sugiere que mi objetivo debería adoptar la forma más simple. Supongámoslo así.

Consideremos el *Dilema de los Padres.* No podemos comunicarnos. Pero cada uno podría o (1) salvar a su propio hijo de algún daño, o (2) salvar al hijo de otro de un daño mayor. Los resultados serían éstos:

		TU: haces (1)	TU: haces (2)
YO	hago (1)	Los hijos de ambos sufren el daño mayor.	El mío no sufre ningún daño. El tuyo sufre ambos daños.
	hago (2)	Mi hijo sufre ambos daños. El tuyo no sufre ninguno.	Los hijos de ambos sufren el daño menor.

Como no podemos comunicarnos, la elección de ninguno de los dos afectará a la del otro. Si el objetivo de cada uno fuera que su propio hijo no sufriera ningún daño, cada cual debería hacer aquí (1) en vez de (2). Cada uno aseguraría así que su hijo sufriera menos. Esto es así, haga lo que haga el otro. Pero si ambos hacen (1) en vez de (2) los hijos de ambos sufrirán más.

Consideremos a continuación los beneficios que debo tratar de procurar a mi hijo. ¿Cuál debería ser aquí mi objetivo? ¿Debo insistir en ser *yo* quien

beneficie a mi hijo si supiera que esto sería peor para el? Alguno respondería «no». Pero esta respuesta tal vez sea demasiado precipitada. Trata los cuidados paternos como meros medios. Creeríamos que son algo más. Podríamos estar de acuerdo en que, con algunas clases de beneficio, mi objetivo adoptaría la forma más simple. Sería simplemente que el resultado fuera mejor para mi hijo. Pero puede haber otros tipos de beneficio, que mi hijo debería recibir *de mí*.

Con ambos tipo de beneficio, podemos afrontar el Dilema de los Padres. Consideremos el *segundo caso*. No podemos comunicarnos. Pero cada uno puede o (1) beneficiar a su propio hijo, o (2) beneficiar al hijo del otro algo más. Los resultados serían éstos:

TU

YO	haces (1)	haces (2)
hago (1)	Lo tercero mejor para nuestros dos hijos.	Lo mejor para el mío, lo peor para el tuyo.
hago (2)	Lo mejor para el tuyo, lo peor para el mío.	Lo segundo mejor para ambos.

Si mi objetivo fuera aquí lograr el mejor resultado para mi hijo, debería hacer una vez más (1) en vez de (2). Lo mismo vale para ti. Pero si ambos hacemos (1) en vez de (2) resultará peor para nuestros dos hijos. Comparemos el *tercer caso*. No podemos comunicar. Pero yo podría, o bien (1) ponerme en condiciones de procurar a mi hijo un beneficio, o bien (2) ponerte en condiciones de beneficiar a tu hijo algo más. Tú tienes las mismas alternativas respecto de mí. Los resultados serían estos:

TU

YO	haces (1)	haces (2)
hago (1)	Cada uno puede procurar a su hijo algún beneficio.	Puedo procurar al mío el máximo beneficio. Tú el mínimo al tuyo.
hago (2)	Puedo procurar al mío el mínimo beneficio. Tú el máximo al tuyo.	Cada uno puede beneficiar a su hijo más.

Si mi objetivo fuera aquí beneficiar yo a mi hijo, debería hacer de nuevo (1) en vez de (2). Y lo mismo vale para ti. Pero si ambos hacemos (1) en vez de (2), cada uno beneficia menos a su hijo. Nótese la diferencia entre estos dos ejemplos. En el Caso Segundo nos importa lo que sucede. El objetivo de cada uno es que el resultado sea mejor para su hijo. Este es un objetivo que el otro puede directamente hacer que se logre. En el Caso Tercero nos importa lo que *hacemos*. Puesto que mi objetivo es beneficiar *yo* a mi hijo, no puedes hacerlo tú por cuenta mía. Pero podrías ponerme en condiciones de hacerlo. Podrías así ayudarme indirectamente a lograr mi objetivo.

Es improbable que ocurran Dilemas del Padre bipersonales. Pero a menudo nos encontramos con versiones multipersonales. Ocurre con frecuencia que si todos, en vez de ninguno, damos prioridad a nuestros propios hijos, o bien resultará peor para todos nuestros hijos, o bien nos colocará a cada uno en condiciones de beneficiarlos menos. Hay pues muchos resultados que beneficiarían a nuestros hijos, ayudemos o no a producirlos. Para cada padre puede ser cierto que sea mejor para sus propios hijos no ayudar. Puede emplear lo que ahorre —ya sea tiempo, dinero o energía— directamente en ellos. Pero que ninguno ayude será para nuestros hijos peor que si todos lo hacen. En otro caso común, cada uno podría, o bien (1) aumentar sus propias ganancias, o bien (2) (imponiéndose restricciones) aumentar las de los demás. En este caso resulta cierto para cada uno que, si hace (1) en vez de (2), podrá beneficiar más a sus hijos. Y así será hagan lo que hagan los demás. Pero si todos hacen (1) en vez de (2), cada uno podrá beneficiar menos a sus hijos. Estos son sólo dos de las posibles formas en que tales casos pueden ocurrir. Pero hay muchas otras.

Observaciones semejantes pueden hacerse respecto de todas las obligaciones semejantes, tales como las que se tiene con alumnos, pacientes, clientes o electores. En todas ellas existen incontables versiones multipersonales de mis tres ejemplos. Son tan comunes y variadas como los dilemas prudenciales Cada Uno-Nosotros. Y como acabamos de ver, a menudo tendrán la misma causa. He aquí otra forma en la que efectivamente sería así. Supongamos que en el caso original son nuestros abogados los que tienen que elegir. Este es el *Dilema del Abogado del Prisionero*. El que ambos abogados den prioridad a sus propios clientes será peor para éstos que el que ninguno lo haga. Todo Dilema prudencial genera así un Dilema moral. Si un grupo se enfrenta con el primero, otro grupo podría, como consecuencia, enfrentarse con el segundo. Podrá ser así si creemos que cada miembro del segundo grupo debe dar prioridad a alguno de los miembros del primero. El problema proviene del hecho mismo de dar prioridad. Da igual que se la dé a uno mismo o a otros.

Todos mis ejemplos incluyen daños y beneficios. Pero el problema puede plantearse respecto de otras partes de la moral del sentido común. Puede plantearse cada vez que esta moral impone deberes diferentes a personas diferentes. Supongamos que cada uno puede, o bien (1) cumplir algu-

nos de sus propios deberes, o bien (2) facilitar a otros un mayor cumplimiento de los suyos. Si todos, en vez de ninguno, dan prioridad a sus propios deberes, cada uno estará en condiciones de cumplir menos de ellos. Los deontologistas pueden enfrentarse con dilemas Cada Uno/Nosotros. Pero no los discutiré aquí.

VII

¿Qué muestran estos casos? La moralidad del sentido común es la teoría moral que la mayoría de nosotros aceptamos. De acuerdo con ella hay ciertas cosas que cada uno de nosotros debemos intentar lograr. Estas son lo que llamo nuestros «objetivos morales». Seguimos con éxito esta teoría moral cuando, de entre los actos disponibles, cada uno realiza el acto que mejor logra sus objetivos morales. En mis casos lo cierto es que, si todos en vez de ninguno siguen con éxito esta teoría, haremos con ello que los objetivos morales de cada uno se logren peor. Nuestra teoría moral es aquí directa y colectivamente autodestructiva. ¿Es eso una objeción?

Comencemos con una cuestión menor. ¿Podríamos revisar nuestra teoría, de forma que no fuera autodestructiva? Si no puede hacerse, la nuestra podría ser la mejor teoría posible. Como creemos en nuestra teoría, deberíamos preguntarnos cuál es la menor de tales revisiones. Deberíamos, por tanto, identificar primero qué parte de nuestra teoría es autodestructiva.

Nos será útil reunir dos distinciones. Una parte de las teorías morales puede cubrir los *actos que tienen éxito* en el supuesto del *cumplimiento pleno.* Llamemos a esta parte la *teoría del acto ideal.* Esta dice qué deberíamos todos intentar hacer en el simple supuesto de que todos lo intentamos y todos tenemos éxito. Llamemos a esto *lo que todos idealmente deberíamos hacer.*

Nótese a continuación que, en mis ejemplos, lo que ocurre es lo siguiente. Si *todos* nosotros seguimos *con éxito* nuestra teoría, ésta será autodestructiva. La que es autodestructiva es nuestra teoría del acto ideal. Si hemos de revisar nuestra teoría, esta es la parte que sin la menor duda debe ser revisada.

La revisión sería ésta. Llamemos M a nuestra teoría. En tales casos todos deberíamos idealmente hacer lo que hiciese que los objetivos M-dados de cada uno se lograsen mejor. Así, en mis Dilemas de los Padres todos deberíamos idealmente hacer (2) en vez de (1). Esto dará un resultado mejor para todos nuestros hijos y nos pondrá a cada uno en condiciones de beneficiar más a los propios hijos.

Llamemos *R* a esta revisión. Nótese en primer lugar que R se aplica únicamente a aquellos casos en los que M es autodestructiva. Si decidimos adoptar R necesitaremos saber cómo pueden reconocerse tales casos. Creo que son muy frecuentes. Pero no tengo espacio para mostrarlo aquí.

Nótese a continuación que R queda restringida a nuestra teoría del acto ideal. No dice qué debemos hacer cuando hay algunos que no siguen R. Ni nos dice cuáles deban ser nuestros objetivos en el caso de que nuestros intentos fallen. Ni nos dice qué disposiciones debamos tener. Como éstas son las cuestiones de la máxima importancia práctica, podría parecer que adoptar R daría casi lo mismo. Pero eso no es probable. Si revisamos esta parte de nuestra teoría, es probable que revisemos el resto. Pongamos por caso un bien público que beneficiaría a nuestros hijos. Un bien de este tipo sería la conservación de un recurso escaso. Supongamos que somos pescadores que intentamos alimentar a nuestros hijos. Nos enfrentamos con existencias decrecientes. Para cada uno es cierto que lo mejor para sus hijos será no restringir sus capturas. Y así será, hagan lo que hagan los demás. Pero para nuestros hijos, que nadie restrinja sus capturas será peor que si todos lo hacen. De acuerdo con R, todos deberíamos idealmente restringir nuestras capturas. Pero si algunos dejan de hacerlo, R deja de aplicarse. Pero sería natural formular esta ulterior pretensión: cada uno debería restringir sus capturas a condición de que un número suficiente de los demás lo hiciese. Tendríamos que decidir cuántos son un número suficiente. Pero decidamos lo que decidamos, adoptar R no habrá dado igual. Dejar de restringir nuestras capturas sería ahora todo lo más un defensivo segundo mejor bien. Considérese a continuación la relación entre actos y disposiciones. Supongamos que cada uno podría, o bien (1) salvar a su propio hijo de un daño menor, o bien (2) salvar al hijo de otro de un daño algo mayor. De acuerdo con R todos deberíamos idealmente hacer (2). ¿Deberíamos *tener la disposición* a hacer (2)? Si los daños menores fuesen en sí mismos grandes, una disposición semejante podría ser incompatible con el amor a nuestros hijos. Esto podría llevarnos a decidir que deberíamos conservar nuestra disposición a hacer (1). Esto implicaría que, en tales casos, nuestros hijos serían más dañados; pero si hemos de amarlos, éste es el precio que ellos deben pagar. Estas observaciones no pueden hacerse cada vez que M sea autodestructiva. Sería posible amar a los propios hijos y contribuir a la mayor parte de los bienes públicos. Ni podrían extenderse a todas las obligaciones similares —tales como las que tenemos para con nuestros alumnos, pacientes, clientes o electores. Es, pues, probable que, si adoptamos R por ello nos veamos inducidos a cambiar nuestro punto de vista sobre algunas disposiciones.

Podemos volver ahora a la cuestión principal. ¿Debemos adoptar R? ¿Constituye una objeción para nuestra teoría moral el que, en algunos casos, sea autodestructiva? Si lo es, R es el remedio obvio, R revisa M sólo allí donde M es autodestructiva. Y la única diferencia es que R no lo es.

Recordemos primeramente que, en esos casos, M es *directamente* autodestructiva. El problema no es que, en nuestros intentos de seguir M, en cierto modo fallemos. Esto podría no ser una objeción. El problema es que todos seguimos M *con éxito.* Cada uno logra realizar, de entre los actos disponibles, aquel que logra mejor sus objetivos M-dados. Esto es lo que con-

vierte a M en auto-destructiva. Y esto sí que parece una objeción. Si existe un supuesto en el que una teoría *no* debería ser autodestructiva, es sin duda el supuesto de que es universalmente seguida.

Recordemos seguidamente que por «objetivos» entiendo objetivos sustantivos. He ignorado nuestro objetivo formal: evitar la actuación incorrecta. Podría parecer que esto elimina la objeción. Fijémonos en los casos en los que, de seguir M, o bien el resultado será peor para todos nuestros hijos, o bien cada uno podrá beneficiar menos a sus hijos. Podríamos decir: «Estos resultados son, ciertamente, desgraciados. Pero ¿cómo podríamos evitarlos? Sólo dejando de dar prioridad a nuestros propios hijos. Pero eso estaría mal. Así, pues, estos casos no arrojan dudas sobre nuestra teoría moral. Ni siquiera para lograr otros de nuestros objetivos morales deberíamos jamás obrar mal».

Estas observaciones son confusas. Ciertamente, en esos casos M no es formalmente autodestructiva. Si seguimos M no estamos haciendo lo que creemos que está mal. Al contrario, creemos que lo que está mal es *no* seguir M. Pero M es sustantivamente autodestructiva. A menos que todos hagamos lo que ahora creemos que está mal, haremos que nuestros objetivos lo que ahora creemos que está mal, haremos que nuestros objetivos M-dados se logren peor. La cuestión es: ¿mostraría esto que estamos equivocados? ¿Acaso debemos hacer lo que *ahora creemos* que está mal? No podemos responder: «No, nunca debemos obrar mal». Si estamos equivocados no estaríamos obrando mal. Ni podemos decir simplemente, «pero, incluso en esos casos, *debemos* dar prioridad a nuestros propios hijos». Esto da precisamente por supuesto que no estamos equivocados. Para defender nuestra teoría debemos ir más allá en nuestras exigencias. Debemos sostener que el que en tales casos sea sustantivamente autodestructiva no constituye una objeción a nuestra teoría.

No constituiría una objeción si, pura y simplemente, no importase que nuestros objetivos M-dados se logren. Pero importa. El sentido en el que importa tal vez no esté claro. Si no hemos actuado mal, quizá no importe moralmente. Pero importa en un sentido que tiene implicaciones morales. ¿Por qué habríamos de intentar lograr nuestros objetivos M-dados? En parte, la razón es que, en este otro sentido, importa que se logren.

Alguien podría decir: «Dicen que M es *autodestructiva*. Así pues, tu objeción ha de apelar a *M*. No puedes apelar a una teoría rival. Y lo acabas de hacer. Cuando afirmas que importa que se logren nuestros objetivos M-dados, afirmas meramente que, si no se logran, el resultado sería peor. Esto da por supuesto el consecuencialismo. De forma que prejuzgas la cuestión.

Pero no es así. Llamemos *neutrales respecto del agente* a nuestros objetivos cuando sean comunes. Otros objetivos son relativos al agente. Cualquier objetivo puede tener que ver, o bien con lo que sucede, o bien con lo que se hace. Así, pues, hay cuatro tipos de objetivos. He aquí algunos ejemplos:

	Interesado en	
	lo que sucede	lo que se hace
neutral respecto del agente	Que los hijos no mueran de hambre.	Que los hijos sean cuidados por sus propios padres.
relativo al agente	Que mis hijos no mueran de hambre.	Que sea yo quien cuide de mis hijos.

Cuando afirmo que importa que nuestros objetivos M-dados se logren, no estoy presuponiendo el consecuencialismo. Algunos de estos objetivos tienen que ver con lo que hacemos. Así el cuidado paternal puede no ser para nosotros un mero medio. Y algo que importa aún más: no estoy presuponiendo el neutralismo respecto del agente. Puesto que nuestra teoría moral es, en su mayor parte, relativa al agente, esto prejuzgaría la cuestión. Pero no hay por qué prejuzgarla.

Hay aquí dos cuestiones de interés. La primera es que no estoy presuponiendo que lo que importa es el logro *de los objetivos M-dados.* Supongamos que puedo, o bien (1) promover mis propios objetivos M-dados, o bien (2) promover más eficazmente los tuyos. De acuerdo con M, debería hacer aquí (1) en vez de (2). Con ello haría que los objetivos M-dados se lograsen peor en su conjunto. Pero con ello no se convierte M en autodestructiva. Yo haré que *mis* objetivos M-dados se logren *mejor.* En mis ejemplos el quid de la cuestión no es que, si todos hacemos (1) en vez de (2), hagamos que nuestros objetivos M-dados se logren peor. El quid es que hacemos que *cada uno de nuestros* objetivos M-dados se logren peor. No obramos peor únicamente en términos de neutralidad respecto del agente, sino en términos de relatividad respecto de éste.

La segunda cuestión es que esto puede importar en términos relativos al agente. Nos servirá de ayuda recordar la prudencia o P. En Dilemas del Prisionero, P es directamente autodestructiva. Si todos más bien que ninguno siguen con éxito P, haremos que el objetivo P-dado de cada uno sea peor alcanzado. Produciremos un resultado peor para cada uno. Si creemos en la prudencia, ¿creeremos que esto importa? ¿O importa sólo que cada cual logre su objetivo formal: evitar la irracionalidad? La respuesta es clara. De acuerdo con la prudencia, actuar racionalmente es un mero medio. Todo lo que importa es el logro de nuestros objetivos sustantivos P-dados. Lo que nos importa es el logro de nuestros objetivos sustantivos P-dados. Lo que nos importa aquí es esto. El logro de estos objetivos importa de un modo relativo al agente. Para considerar que constituye una objeción el que nuestra prudencia sea autodestructiva, no necesitamos apelar a su formulación

en términos neutrales respecto del agente: el Utilitarismo. La prudencia no es una teoría moral. Pero la comparación muestra que, al discutir la moralidad del sentido común, no tenemos por qué prejuzgar la cuestión. Si importa que se logren nuestros objetivos M-dados, también puede importar que se logren de forma relativa al agente.

¿Importa esto? Nótese que no pregunto si es lo único que importa. No sugiero que el logro de nuestro objetivo formal —evitar obrar mal— sea un mero medio. Aunque los consecuencialistas lo presuponen, no es lo que creemos la mayoría de nosotros. Podemos incluso creer que el logro de nuestro objetivo formal siempre es lo más importante. Pero aquí resulta irrelevante. Lo que preguntamos es si el hecho de ser sustantivamente autodestructiva arroja dudas sobre M. ¿Podría esto mostrar que, en tales casos, M es incorrecta? Tal vez sea verdad que lo que más importa es que evitemos obrar mal. Pero esta verdad no puede mostrar que M sea correcta. No puede ayudarnos a decidir lo que está mal.

¿Podemos afirmar que nuestro objetivo formal es todo cuanto importa? Si así fuera, mis ejemplos no mostrarían nada. Podemos decir, «Ser sustantivamente autodestructiva es, en el caso de la moralidad del sentido común, no ser autodestructiva». ¿Podemos defender nuestra teoría moral de este modo? En el caso de algunos objetivos M-dados tal vez podamos. Consideremos promesas triviales. Podríamos creer tanto que debemos tratar de cumplirlas, cuanto que no importa si, sin culpa por nuestra parte, no las cumplimos. Pero no es esto lo que creemos respecto de todos nuestros objetivos M-dados. Importa que nuestros hijos sufran daño o que podamos beneficiarlos menos.

Recordemos finalmente que, en mis ejemplos, M es colectiva, *pero no individualmente* autodestructiva. ¿Podría servir esto como defensa?

Esta es la cuestión central que he planteado. Es precisamente por tener éxito individualmente, por lo que M es aquí *directamente* autodestructiva en el nivel colectivo. ¿Por qué es verdad que, si todos hacemos (1) en vez de (2), seguimos *con éxito* M? Porque *cada uno,* de entre los actos disponibles, realiza aquel que *mejor* logra sus objetivos M-dados. ¿Será tal vez que no constituye una objeción el que con ello *nosotros* hacemos que los objetivos M-dados de cada uno se logren *peor*?

Nos servirá de ayuda una vez más recordar la prudencia. En los Dilemas del Prisionero, la prudencia es colectivamente autodestructiva. Si estuviésemos eligiendo un código colectivo, uno que todos seguiremos, la prudencia nos sugeriría que la rechazáramos. Sería prudente votar contra la prudencia. Pero quienes creen en la prudencia, pueden pensar que esto es irrelevante. Pueden decir: «La prudencia no pretende ser un código colectivo. Ser colectivamente autodestructiva no es, en el caso de la prudencia, ser autodestructiva».

¿Podemos defender nuestra teoría moral de este modo? Depende de nuestra concepción de la naturaleza de la moralidad. En la mayoría de las concepciones, la respuesta es «No». Pero debo dejar aquí la cuestión abierta.

Título original: *Prudence, Morality and Prisioner's Dilemma.*
Tomado de: *Proceedings of the British Academy,* London, vol LXV (1979), Oxford University Press, pp. 539-564.
Traducción: José M.ª VEGAS MOLLÁ