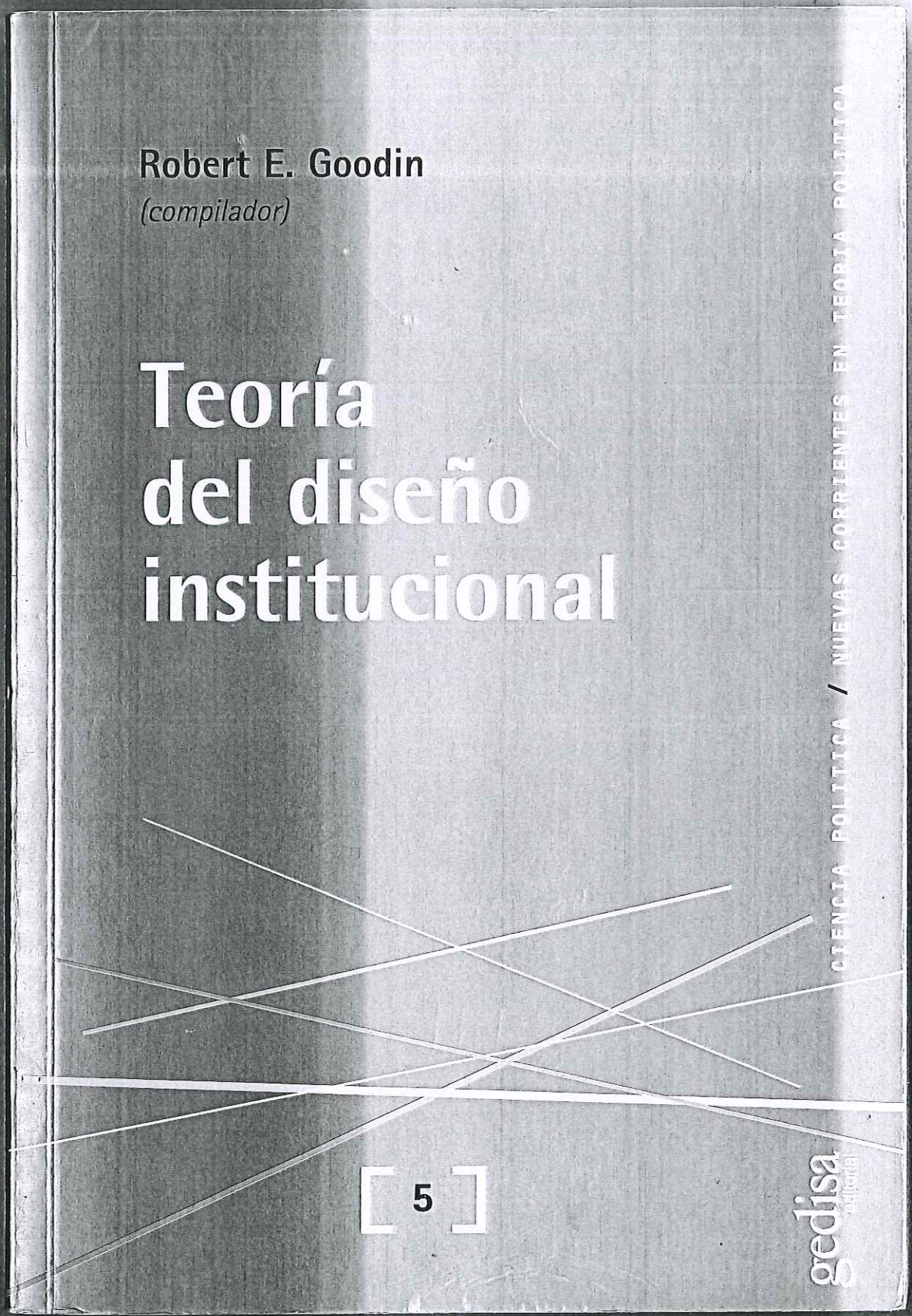
Robert E. Goodin
(compilador)
Teoría del diseño institucional
CIENCIA POLITICA / NUEVAS CORRIENTES EN TEORIA POLITICA
5
gedisa

2

El diseño institucional y la elección racional

PHILIP PETTIT

En este capítulo se considera al diseño institucional desde la perspectiva de la teoría de la elección racional. Me propongo reconocer los principios del diseño institucional –de la regulación social, en sentido amplio– que resultan válidos dentro de la mejor interpretación del enfoque de la elección racional.

El proyecto de analizar al diseño institucional desde la perspectiva de la teoría de la elección racional resultará atractivo para aquellos que están dispuestos a aceptarla. Sin embargo, ¿resulta acaso probable que tenga algún interés para quienes la rechazan? Considero que sí. La teoría de la elección racional equivale a «las ciencias sociales por medios economicistas» y ha tenido una marcada influencia entre los que estudian políticas y los que las proyectan, quienes con frecuencia tienen una formación dentro de la disciplina de la economía. Sólo por esa razón resulta de interés determinar cuál es la enseñanza que nos deja –o que debería dejarnos– esta teoría con respecto al diseño institucional.

Existe, asimismo, una consideración adicional. Es habitual que los teóricos normativos califiquen los supuestos de la teoría de la elección racional de simplistas y sectarios, y que supongan que las lecciones de la teoría respecto al di-

seño institucional son relativamente directas. Pero la realidad es que el enfoque de la elección racional es un punto de vista relativamente sofisticado sobre la conducta humana (al menos si se hace una descripción indulgente) y que ofrece una perspectiva minuciosa del diseño institucional adecuada para muchas de las revelaciones que ofrecen los enfoques más sociológicos.[1] La teoría nos proporciona razones para admitir determinados principios del diseño institucional, que se enumeran en la sección final, los cuales probablemente resulten interesantes para teóricos de diversas extracciones.

2.1. El diseño institucional

Cuando hablo de *diseño* institucional, no necesariamente me refiero a la creación de acuerdos sociales completamente nuevos. La frase incluye ese caso, por cierto, pero mi intención es que sea también aplicable al proyecto más habitual de examinar acuerdos existentes para determinar si resultan satisfactorios y modificarlos cuando fuera necesario; el proyecto de replantear y reformular (quizás con relativa modestia) y no el proyecto de darles su forma inicial. Quizás sería mejor hablar de intervención institucional y no de diseño institucional; lo haría con gusto, si se determinara que el término resulta preferible.

Cuando hablo de diseño *institucional,* no hago alusión al diseño y rediseño de estructuras meramente formales (por ejemplo, estructuras constitucionales para la organización del Parlamento o de los tribunales). Utilizo esta frase para incluir intervenciones en todos los acuerdos que coordinan la conducta de los individuos dentro de la sociedad. Tales acuerdos incluyen los procedimientos establecidos constitucional o legalmente, pero también abarcan cuestiones que están apenas sujetas a normas y convenciones conscientes o que están fijadas únicamente por presiones y perspectivas tácitas o registradas acaso ocasionalmente.

¿Cuál es el propósito de la teoría del diseño institucional? ¿Quiénes son los diseñadores potenciales a los cuales está dirigida? Desde mi concepción, la respuesta es la siguiente: todos aquellos que tienen un interés en la manera en la que está organizada la vida social (como debería tenerlo todo ciudadano democrático) y se encuentran en posición de proponer cambios diseñados para lograr una reforma (como asimismo debería estarlo todo ciudadano democrático). A todas luces, existe una variedad de problemas graves en toda sociedad y todos aquellos que se ven inclinados a reflexionar acerca de ellos deben tener un interés en el diseño institucional; es decir que su ejercicio no tiene un carácter elitista.

Con esto concluimos esta introducción informal. Desde un punto de vista más analítico, observo tres supuestos que están vinculados con el proyecto del diseño institucional.

1. La conducta de los individuos dentro de la sociedad –su conducta individual, su conducta en tanto agentes de cuerpos corporativos y su conducta en funciones oficiales– es sensible a las oportunidades y los incentivos que están disponibles por efecto de su situación social, al igual que resulta sensible a otros factores (por ejemplo, los valores, representaciones y modalidades discursivas que heredan, en gran medida, de su trasfondo social).
2. Las oportunidades e incentivos asociados a determinada situación social, a menudo, pueden ser modificados –es decir, pueden ser institucionalmente diseñados– de manera tal que se produzca una variación en la conducta agregada de los individuos. En particular, a menudo pueden modificarse con un efecto más determinado o inmediato sobre la conducta agregada del que pueden lograr los otros factores relacionados con la conducta humana.
3. Existen algunos criterios de evaluación, de aceptación relativamente amplia, para determinar que ciertos patrones agregados de conducta resultan más deseables que otros y que, por consiguiente, puede resultar atractivo promoverlos mediante la modificación de las oportunidades o incentivos para los agentes pertinentes; es decir, intentar un diseño institucional a fin de que se establezcan tales patrones.

El primer supuesto postula algunos motores de la conducta humana, como podríamos denominarlos, e identifica algunos de ellos –las oportunidades y los incentivos– que están asociados, de manera genérica, con el entorno social actual. No significa que éstos sean los únicos motores que existen. Por el contrario, nos señala otros factores: los valores adquiridos que tienen los agentes; sus representaciones adquiridas con respecto a las

1. Comparto, definitivamente, el espíritu del siguiente comentario de Ayres y Braithwaite (1992, p. 51): «Gran parte de las ciencias sociales contemporáneas está en un punto muerto entre las teorías que suponen la racionalidad económica de los actores y las teorías que objetan que la acción tiene variadas motivaciones tales como el deseo de cumplir las normas, el de mantener un sentido de identidad, el de hacer el bien o, simplemente, el de actuar conforme una secuencia conductista convertida en hábito. Consideramos que es más probable descubrir ideas sólidas de política cuando nos concentramos en las áreas de convergencia entre los análisis que se basan en el *Homo economicus* y aquellos que se basan en el *Homo sociologicus*».

situaciones en las cuales se encuentran; las maneras adquiridas de debatir acerca de ellas.

El segundo supuesto afirma que los motores identificados no se limitan a serlo, sino que también sirven como palancas (Brennan y Pettit, 1991). Son factores que pueden modificarse, a través de un ejercicio de diseño institucional, a los fines de variar la conducta agregada de los individuos. Nótese que otros motores de la conducta individual, tales como los valores, las representaciones y los discursos, que ya mencionamos, pueden no servir como palancas institucionales de este tipo. Es posible que no sean fáciles de modificar y que un cambio genere únicamente efectos a largo plazo relativamente indeterminados.[2]

El tercer supuesto sugiere que resulta válido examinar las posibilidades y los efectos de modificar determinadas oportunidades e incentivos; es decir, que el diseño institucional tiene sentido. Si no existiera absolutamente ningún consenso acerca de las cuestiones vinculadas con el valor, el ejercicio difícilmente atraería el interés general. Podría suscitar el interés de todos, pero lo haría por una razón distinta en cada caso, porque demostraría cómo realizar modificaciones que se adecuaran a los gustos particulares de una persona. No obstante, considero que en las sociedades como las nuestras existe una buena medida de acuerdo evaluativo, aunque también exista una gran diversidad. Considero, por ejemplo, que todos estaríamos de acuerdo en cuestiones como las siguientes: que los políticos deberían ser honestos; que los policías no deberían inventar pruebas para incriminar a una persona; que los jurados deberían guiarse por su conciencia; que el gerente de una fábrica debería prestar atención a la seguridad industrial; que nadie debería hacer daño a un inocente. Aunque existan diferentes filosofías políticas en nuestras sociedades, todas convergen en una serie de recomendaciones de este tipo.

¿Qué medidas se consideran, dentro del diseño institucional, como instrumentos para modificar las oportunidades e incentivos de los agentes individuales y, por consiguiente, su conducta agregada? (Brennan y Pettit, 1993). Los instrumentos más obvios son los que se describen como sanciones, en sentido amplio. Las sanciones operan sobre el conjunto de opciones que tiene un agente, haciendo que algunas de ellas resulten más –o menos– atractivas de lo que serían de no existir tales sanciones; es decir, afectan los incentivos pertinentes. Las sanciones pueden adoptar la forma positiva o la

2. Puede incluso darse el caso de que los factores involucrados sean susceptibles de ser alterados pero no a través del diseño positivo o, al menos, no de manera confiable. Véase el argumento acerca de las estrategias motivadoras en la sección 2.3.

negativa, como recompensas o como penas. La sanción positiva recompensa al agente por elegir de manera adecuada; la negativa lo castiga por hacerlo incorrectamente.[3]

Resulta claro que se trata de sanciones cuando podemos identificar una agencia deliberada sancionadora. En este caso, nos concentramos en la situación que se produciría en ausencia de la agencia y consideramos que sus respuestas a los actos de las partes pertinentes son intentos de ejercer su influencia, al imponer recompensas y penas apropiadas. No obstante, las sanciones (incluso del tipo previsto por el diseño institucional) no necesariamente involucran la presencia de una autoridad sancionadora de este tipo. Supongamos que existieran determinadas recompensas y penalizaciones que se ofrecieran a los agentes en un tipo de situación y que hicieran más probable que realicen una determinada elección que si éstas no existieran. En una situación así, podríamos decir que están en funcionamiento sanciones. No es relevante que no haya un único agente o agencia que imponga las recompensas y los castigos. No es relevante que éstos no sean impuestos intencionalmente. No es relevante, incluso, que no sean impuestos por otros agentes; pueden, incluso, ser producto de causas naturales.

Las sanciones, en sentido amplio, normalmente son reconocidas por las partes a quienes afectan y pueden estar interiorizadas en su proceso de deliberación. Las partes llegan a la conclusión de que la opción favorecida por las sanciones es la que deben elegir, dada la recompensa que conllevan o el castigo que evitan. No obstante, las sanciones también pueden funcionar aunque no se las tenga en cuenta de esta manera e, incluso, aunque no surjan nunca en la conciencia de los agentes a quienes afectan. Supongamos que se dé una situación en la cual los agentes ya tengan razones suficientes para elegir una determinada opción pero que cualquier agente instrumental establezca sanciones que favorezcan esa opción. Aunque los agentes en cuestión no estén al tanto de las sanciones, éstas funcionarán de todas maneras haciendo más infalible la elección deseada que en el caso anterior. Su efecto puede ser que, si las razones de los agentes los desvían de la opción deseada y éstos comienzan a elegir otras opciones, generalmente se vayan dando cuenta de la recompensa que pierden o de la pena que se les aplica y regresen, por lo tanto, a su opción original. Así, las sanciones pueden funcionar haciendo que la elección de una op-

3. Casi invariablemente se da por sentado que las sanciones se basan en el interés egoísta, de manera tal que es la persona misma la que recibe el castigo o la recompensa y no un tercero, ni siquiera alguien que resulte importante para ella. Coincido con este supuesto. La teoría de la elección racional se combina con la intuición moral para volverla convincente.

ción determinada resulte más segura, sin figurar explícitamente en la deliberación del agente. Pueden servir para reforzar una determinada forma de conducta, incluso aunque no contribuyan a producirla.

El concepto de sanción nos resulta familiar a todos en la vida social. Pero el diseñador institucional cuenta con un segundo tipo de control, que no está tan ampliamente reconocido. Éste es el filtro o selección. Las sanciones que puede implantar un diseñador institucional toman como dados a los agentes y las opciones e intentan influir sobre la elección al modificar el interés relativo de estas últimas para los agentes dados, es decir, afectando sus incentivos. Un filtro concebido por un diseñador opera, por el contrario, sobre el conjunto de agentes u opciones. Su propósito es asegurar que determinados agentes puedan realizar determinadas elecciones, y no otras, o que en una determinada elección estén disponibles algunas opciones y no otras. En otras palabras, están diseñados para influir sobre las oportunidades y no sobre los incentivos. Los filtros pueden ser positivos o negativos de la misma manera que las sanciones (recompensas o penas). Pueden eliminar determinados agentes u opciones o bien incluir nuevos agentes u opciones, lo cual es quizás más sorprendente. Pueden otorgar poder a individuos que previamente no estaban involucrados en la situación en cuestión, dándoles una oportunidad para actuar que no tenían anteriormente; o bien pueden otorgar poder a individuos que ya se encontraban involucrados, ofreciéndoles una opción nueva en su lista de alternativas.[4]

En el caso del diseño institucional ideal, los filtros que operan sobre los individuos tienen el efecto de atraer hacia determinadas tareas a aquellos individuos que tienen más probabilidades de conducirse de una manera que resulta socialmente valiosa (es posible que las tengan de manera inherente o dentro del contexto de determinadas sanciones). Los procedimientos de designación, recusación y llamado a concurso, así como los requisitos y las restricciones para ser nombrado en un puesto público, son ejemplos del tipo de filtros que un diseñador institucional intenta controlar. Consideremos, por ejemplo, el procedimiento que sujeta a recusación a los miembros de un comité para garantizar que no tengan intereses personales –el interés de un amigo o colega– en el resultado que se proponen dictaminar. O bien, el requisito de que los miembros del comité incluyan representantes de determinados grupos; que sean aprobados por una autoridad independiente; o que las partes interesadas puedan recusarlos. Todas estas medidas representan obvios mecanismos de filtro.

4. Agradezco a John Braithwaite por recordarme la importancia de los filtros positivos.

Sin embargo, los filtros pueden aplicarse tanto sobre las opciones como sobre los agentes, incluyéndolos o excluyéndolos de la lista de alternativas disponibles. Es de esperar que encontremos este tipo de mecanismo de filtro cuando la disponibilidad de una opción depende (o es posible que dependa) del acceso a determinados recursos, sociales o físicos, determinados por factores que se encuentran bajo el control del diseñador institucional. El hecho de que una investigadora se dedique a un determinado proyecto depende a menudo de que exista una organización dispuesta a suministrarle los fondos indispensables. Por lo tanto, un proyecto puede verse trabado por una negativa a financiarlo o verse facilitado por la disponibilidad de fondos. Asimismo, el proyecto al que se aboca una investigadora depende habitualmente de que la organización a la que pertenece esté dispuesta a autorizarlo (o es posible hacer que dependa de ello) y, por consiguiente, también es posible bloquear o facilitar el proyecto por medios jurídicos (por oposición a los económicos). No es sorprendente, entonces, que las personas intenten que determinado tipo o calidad de proyectos se vean favorecidos a través del diseño institucional, y que sólo los proyectos que cumplen determinados criterios éticos sean posibles.

Para concluir esta sucinta argumentación acerca del diseño institucional, puede resultar útil indicar la gama de casos con respecto a los cuales resulta potencialmente pertinente. El espectro es enorme, pero puede ensayarse una taxonomía de los casos en dos dimensiones, al menos por razones mnemotécnicas. El objetivo del diseño institucional puede ser la prevención del daño o la promoción del bien. La distinción es intuitiva –como así también su relevancia– aunque plantee algunos problemas analíticos. Por otra parte, el diseño institucional puede proponerse el control de los agentes normales individuales o corporativos –en particular, los poderosos y los peligrosos– o de aquellos que son depositarios de una cierta confianza social: los individuos y cuerpos colegiados que asumen determinados deberes públicos. La taxonomía se representaría de la siguiente manera:

	Prevención del daño	*Promoción del beneficio*
Agentes privados	Sección 2.1.1.	Sección 2.1.2.
Agentes públicos	Sección 2.1.3.	Sección 2.1.4.

La mejor manera de señalar el alcance potencial del diseño institucional consistiría en tomar cada una de las categorías que nos brinda esta cla-

sificación –que representa la clase de resultados que desea facilitar el diseñador institucional– y ofrecer un ejemplo. En cada caso ilustrado, el diseñador institucional se propone identificar el tipo de mecanismos de filtro o de sanción que resulta posible implementar, a fin de mejorar el nivel actual de desempeño. Es posible que esta manera de enfocar los casos resulte algo extraña, ya que es necesario seleccionar unos pocos ejemplos de una gama de posibilidades muy amplia. Indudablemente, también puede parecer ingenuo, ya que muchos de los casos son tan evidentes que mencionarlos parece mortificante. Pero la consideración de estos casos nos asegura, por lo menos, un útil sentido concreto del tema en cuestión.

2.1.1. Prevención del daño por parte de agentes privados

El daño al que nos referimos aquí incluye el perjuicio que alguien como usted o yo podemos causar, del tipo normalmente prohibido por el derecho penal, pero también el daño que pueden producir aquellos que se encuentran en puestos de poder e influencia: la difusión de información errónea por parte de los medios de comunicación; la negligencia en cuestiones de seguridad industrial por parte de la gerencia de una fábrica; el daño ecológico que causan diversas empresas, etcétera. Uno de los principales objetivos del diseño institucional, en toda sociedad, debe ser la minimización de este tipo de conducta dañosa. Esto se logra estableciendo filtros y sanciones que reduzcan la capacidad de causar un potencial daño o que aumenten el poder de la potencial víctima (quizás, por ejemplo, a través de la inclusión de determinadas opciones a través de filtros positivos).

2.1.2. Promoción del beneficio por parte de agentes privados

En este caso, un ejemplo clásico puede ser la provisión de sangre para transfusiones que comentara Richard Titmuss (1971), pero existen otros numerosos ejemplos. Éstos incluyen el voto a conciencia, al menos en ciertos aspectos (Brennan y Pettit, 1990); la contribución a causas valiosas, vinculadas a la caridad o la cultura; el establecimiento de fideicomisos y fundaciones para la promoción de la ciencia, el arte o la tolerancia religiosa; la apertura al público de colecciones de arte privadas. Sería de esperarse que el diseñador institucional desee facilitar tales beneficios, sea cual fuere el contexto en cuanto a filosofía política.

2.1.3. Prevención del daño por parte de agentes públicos

Es un lugar común afirmar que algunos de quienes detentan un cargo público (en particular, los políticos, los jueces y las fuerzas armadas y de seguridad) pueden ocasionar, en muchos casos, un gran daño al conjunto de la sociedad a cuyo servicio se encuentran. Su poder les da la oportunidad de favorecer diversos intereses privados, lo que muchas veces conlleva severos costos para la comunidad y los expone a una tentación y presión considerables. Dentro del diseño institucional ideal, buscaremos medidas adecuadas –un patrón apropiado de filtros y sanciones– a fin de promover en tales autoridades una conducta correcta desde el punto de vista procedimental guiada por el espíritu del servicio público.

2.1.4. Promoción del beneficio por parte de los agentes públicos

Todo mandatario público tiene una tarea que cumplir, tanto cuando el mandato en cuestión implica el tipo de poder al que hicimos referencia como cuando no es así. Asimismo, uno de los axiomas del diseño institucional es la creencia de que esto debe hacerse correctamente, de que los políticos, los jueces y las fuerzas militares y de seguridad deben llevar a buen término sus funciones oficiales y que también los funcionarios menos poderosos deben desempeñarse de una manera satisfactoria. La categoría de los funcionarios incluye a quienes ejercen muchas de las profesiones vinculadas con el cuidado de la salud, con la enseñanza y con la investigación y, por supuesto, a los miembros de la burocracia oficial. No obstante, debemos tener en cuenta que también incluye a muchos de quienes detentan responsabilidades temporarias: los jurados, que sólo prestan servicios en un único juicio; los miembros de un comité investigador de una transgresión específica; en fin, aquellos que prestan servicios en un comité al que se le ha asignado un mandato específico para una designación, promoción o resolución.

2.2. La teoría de la elección racional

Para utilizar una frase que ya he mencionado, la teoría de la elección racional puede describirse como «las ciencias sociales por medios economicistas» (Elster, 1986a). Se resume como un intento de lograr una explicación en términos economicistas no sólo del comportamiento del mercado, sino

también de la conducta externa al mercado. La idea que guía esta perspectiva es que, si la economía sirve para explicar la manera en la que se comportan los agentes en contextos más o menos asimilables a un mercado, debería servirnos igualmente para la explicación de su conducta en otros terrenos. Este enfoque ha sido examinado con diferentes nombres. La explicación que propone la elección racional para la conducta política se ha presentado con el nombre de teoría de la elección pública, por ejemplo, y la explicación que propone la teoría de la elección racional para la interacción social, con el de teoría de los intercambios (McLean, 1987; Heath, 1976). Pero se mantiene, de manera relativamente coherente, la idea de poner el método de las ciencias económicas al servicio de una explicación para la conducta económica tanto como aquella que no lo es.

Desde mi punto de vista, la teoría de la elección racional se diferencia de las teorías abstractas de la racionalidad práctica desarrollada en ciertas áreas de investigación dominadas por una multitud de economistas: la teoría de la decisión, la teoría de los juegos y la teoría de la elección social (Hargreaves-Heap et al., 1992). Quienes proponen la teoría de la decisión intentan enunciar, en abstracto, qué significa que un agente sea completamente racional. Así, los teóricos dentro de la línea de Bayes explican el concepto de agente racional como aquel que maximiza la utilidad esperada (Eells, 1982). Quienes suscriben la teoría de los juegos intentan identificar soluciones, de existir alguna, que se espera que alcancen los agentes racionales en distintas situaciones de decisión interdependiente o juegos, para emplear esa desafortunada denominación (Luce y Raiffa, 1957). Los teóricos de la elección social se proponen identificar el orden de las preferencias y la elección que debería adoptar racionalmente un grupo –de existir una opción racionalmente interesante– dados diversos ordenamientos de preferencias para los individuos que lo componen (Sen, 1970).

La teoría de la elección racional puede basarse, en diferentes aspectos y de distintas maneras, sobre estos tres cuerpos de teoría más abstracta. Pero, en sí misma, constituye un ejercicio mucho más concreto y problemático. Se ocupa de personas reales dentro del mundo real y no de agentes idealmente racionales; se propone explicar y predecir la conducta de estas personas, y no impartir lecciones sobre qué resulta racionalmente normativo para ellas (véase Pettit, 1993a, cap. 5; 1993b).

Sin embargo, si la teoría de la elección racional constituye un sistema de explicación y predicción, ¿cuáles son sus postulados centrales? De acuerdo con Michael Taylor (1988), presentaré un breve comentario de

John Harsanyi (1969, p. 524) como una acertada expresión del contenido de la teoría: «La conducta de los individuos puede explicarse, en gran medida, en función de dos intereses dominantes: el beneficio económico y la aceptación social» (véase también Becker, 1976, cap. 1). Cuando Harsanyi hace alusión a la conducta en estos términos, supone que debe explicársela como racionalmente determinada o restringida –dadas las creencias del agente– por esas preocupaciones guiadas por el interés egoísta. Lo que nos dice es, entonces, lo siguiente: en primer lugar, que la teoría de la elección racional infiere un sentido racional de la conducta de las personas; en segundo lugar, que lo hace por referencia al interés egoísta; en tercer lugar, que el interés egoísta invocado puede ser un interés en el beneficio económico o en la aceptación social. Además, plantea esta afirmación teniendo en cuenta una salvedad: que el sentido inferido no es, necesariamente, completo y que los términos propuestos pueden explicar en gran medida la conducta de los individuos, pero no en su totalidad.

La primera línea de razonamiento no atraerá demasiado disenso. Si fuera necesaria una ulterior explicación de lo que significa «racional», probablemente se recurriría a la teoría de la decisión para completarla. De no ser así, puede ofrecérsenos en su lugar un sentido informal del tipo de racionalidad (racionalidad en el sentido de Hume) que plantea la teoría de la decisión. El significado de esa racionalidad es, aproximadamente, el siguiente: la elección de un agente resulta racional siempre que promueva la satisfacción de sus deseos mejor que cualquiera de las alternativas, según su manera de ver las cosas. Es decir, sólo en tanto satisfaga sus deseos de acuerdo con sus creencias. Cabe señalar que esta concepción de la racionalidad significa que los teóricos de la elección racional deben formular supuestos acerca de las creencias que comparten los agentes cuya conducta se proponen explicar. Tales supuestos no son dictados, en sí mismos, por la teoría de la elección racional –pueden ser el resultado de la influencia de otras teorías, incluso de carácter muy distinto– aunque se planteará la cuestión de si resultan racionales en la acepción de racionalidad que se aplica a las creencias.

La segunda línea que plantea la fórmula de Harsanyi atraerá algunas objeciones. Se dirá, como lo hacen muchos economistas, que la explicación de la elección racional no necesita postular el interés egoísta y que es posible ser neutral con respecto a qué tipo de deseos mueven a los agentes, como en el caso de la teoría de la decisión, y guiarnos, simplemente, por la idea de que determinados deseos o preferencias funcionan de modo tal que la conducta representa un intento racional de satisfacerlos. Es posible acompañar el espíritu de la teoría de las preferencias reveladas, se-

gún la cual la economía puede identificar en la conducta de los agentes las mismas preferencias o deseos que explican tal conducta.

Hay tres puntos que deseo aclarar en defensa de la invocación que hace Harsanyi del interés egoísta. La primera idea es que, cualquiera que sea la teoría, la práctica de los economistas y de los teóricos de la elección racional consiste en apelar principalmente a los deseos del interés egoísta al intentar explicar y predecir la conducta humana.

La segunda señala que, a menos que los economistas y los teóricos de la elección racional se comprometan con algún postulado sustantivo acerca del tipo de deseos a cuyo servicio se encuentra generalmente la conducta humana, es probable que su proyecto pierda relevancia. Siempre será posible encontrar deseos y creencias tales que pueda considerarse que un determinado fragmento de conducta está al servicio de esos deseos de acuerdo con esas creencias. Efectivamente, de no existir otras restricciones a los deseos y creencias que se pueden invocar, siempre se podrá encontrar una variedad indefinida de conjuntos de creencias y deseos de este tipo (Davidson, 1984). Por lo tanto, el proyecto explicativo y predictivo de la teoría de la elección racional se encuentra en una posición vulnerable ante la ausencia de un postulado sustantivo sobre los deseos de los individuos.[5]

La tercera idea que deseo dejar establecida, en apoyo del postulado del interés egoísta que introduce Harsanyi, recurre –de manera libre y poco ortodoxa– a una observación de Amartya Sen (1982, parte 1). El autor señala que, aunque los economistas afirmen a menudo que siguen la línea de la preferencia revelada, en la práctica tratan las preferencias que atribuyen a los individuos como si estuvieran basadas en el interés egoísta, en la medida en que suponen que para un individuo siempre será mejor (en un sentido aproximadamente utilitarista de producir un aumento del bienestar) que éstas se vean satisfechas, cualquiera que sea la preferencia asignada. Este supuesto no puede ser avalado en el caso de los deseos que no responden al bien propio, como lo demuestra el antiguo relato del niño que encuentra dos manzanas y le da la más pequeña a su amigo. El amigo se queja, diciendo que, si las hubiera encontrado él, le hubiese dado la manzana más grande al otro. Éste le responde que si es verdad que, en su lugar, su amigo hubiese renunciado a la manzana más grande, si realmente

5. Los supuestos arquetípicos de la teoría del consumidor –que no llegan a constituir un modelo propio de motivación a partir del interés egoísta, aunque se ajusten con naturalidad al modelo– representan ya una cierta restricción a los deseos de las personas. Sugieren, por ejemplo, que las curvas de la demanda tienen pendiente negativa, poco pronunciada y cóncava.

es tal su preferencia, no tiene nada de qué quejarse ya que ésta ha sido completamente satisfecha al recibir, de hecho, la manzana más pequeña.

Supongamos, entonces, que podamos caracterizar a la teoría de la elección racional como la idea de que la conducta de las personas está, en gran medida, determinada o restringida racionalmente por deseos guiados por el interés egoísta. La última cuestión que se plantea es si resulta justo afirmar que los deseos basados en el interés egoísta se dividen en deseos de beneficio económico y deseos de aceptación social.

La categoría del beneficio económico incluye, por cierto, la obtención de todas las formas de moneda y todas las formas de bienes comerciales, pero se extiende más allá. Debe incluir, además, el goce de los servicios que brindan los demás, aunque no sean comerciales; el goce de los bienes públicos, que implican que si alguien puede usarlos todos los demás también pueden; y el goce de bienes materiales que nadie suministra, tales como un clima agradable y un paisaje bello. ¿Qué tienen todos en común? La oferta o accesibilidad de los bienes depende de la acción intencional, propia o ajena. Los bienes son, claramente, dependientes de una acción.

Normalmente, es posible suponer, dentro de la tradición de la elección racional, que los únicos bienes que podemos desear para nosotros mismos son los que dependen de una acción, de manera que el supuesto de que las personas son motivadas por el interés egoísta equivale al supuesto de que buscan su propio beneficio económico (Holmes, 1990). Pero la categoría de la aceptación social nos señala un tipo diferente de bienes: bienes que dependen de la actitud, en particular bienes que dependen de la actitud de los demás. Los bienes que dependen de una acción se obtienen en virtud de lo que hace una persona o un tercero. Los bienes que dependen de la actitud se logran gracias a lo que piensa la persona o los terceros.[6] Entre éstos se incluye el bien de la autoestima, del cual gozo en la medida en que llegue a pensar bien de mí mismo. También, y esto es más relevante para la aceptación social, incluyen bienes como la estima, la gratitud, el afecto de terceros, que se tienen en la medida en que los demás lleguen a pensar bien o con cariño de uno.

La idea de que la aceptación social motiva a los agentes interesados al mismo nivel que el beneficio económico no es nueva (Lovejoy, 1961, Lección 5). A pesar de que se lo asocia con el enfoque de la elección racional y económica, Adam Smith es uno de sus defensores más elocuen-

6. Según ha llegado a convencerme Rae Langton, la dicotomía más clara se daría entre los bienes que dependen de la actitud y los bienes –del tipo económico– que son independientes.

tes. De hecho, ha sugerido que a menudo es debido al deseo de aceptación social que las personas buscan la ganancia económica. «La naturaleza, al formar al hombre para la sociedad, le otorgó un deseo original de agradar a sus hermanos y una aversión original a ofenderlos. Le enseñó a sentir placer por su opinión favorable y aflicción por la desfavorable. Hizo que su aprobación, en sí misma, fuera extremadamente halagadora y agradable para él y que su desaprobación resultara asaz mortificante y ultrajante» (Smith, 1982, p. 115).

Propongo que, como lo hace Harsanyi, consideremos que la teoría de la elección racional concibe al deseo egoísta de lograr la aceptación social a la par del deseo egoísta de beneficio económico. Esto tiene la desventaja de volver menos exacta la teoría, en especial debido a que no se asigna una valuación entre los dos tipos de deseos interesados. Pero esa desventaja se compensa por la abrumadora verosimilitud del supuesto de que la aceptación social está vinculada al interés egoísta de los individuos al igual que la ganancia económica.[7]

Para recapitular, entonces, conforme la posición de John Harsanyi, hemos caracterizado la teoría de la elección racional como la teoría de que la conducta de los individuos puede explicarse como determinada o restringida racionalmente por ciertas preocupaciones dictadas por el interés egoísta, en particular la preocupación por el beneficio económico y la aceptación social. Ahora debemos ocuparnos de la restricción general implícita en esta caracterización que señala que la conducta humana puede explicarse, al menos parcialmente, de esta manera. ¿Qué significaría afirmar que los intereses explican «en gran medida» la conducta de los individuos, como lo enuncia Harsanyi, por oposición a una explicación más

7. Una objeción. Claramente, incorporar a la aceptación social no plantea un problema en la medida en que aceptemos que los agentes movidos por su interés egoísta la desean de manera relativamente independiente de su deseo por la ganancia económica, es decir, si ambos bienes son deseados por sí mismos. Sin embargo, supongamos que alguno de los dos resulta deseable únicamente porque promete al individuo una cuota mayor del otro. Supongamos, por ejemplo, como pueden sugerir muchos teóricos de la elección racional, que la aceptación social se desea únicamente debido al hecho de que nuestras posibilidades de gozar de determinados beneficios económicos que dependen de terceros (quizás en un futuro indefinido) aumentan si somos socialmente aceptados por ellos. ¿No significa esto que, si observamos al beneficio económico y a la aceptación social que asegura una opción separadamente, podríamos estar contando dos veces lo mismo? ¿No podríamos estar contando dos veces el beneficio económico asociado con la aceptación? No creo que sea necesario que nos preocupemos por esto. Cualesquiera que sean las perspectivas de beneficio económico asociadas con una opción en virtud de su promesa de una mayor aceptación social, resulta improbable que las incluyamos en nuestra sumatoria independientemente de nuestra consideración de la promesa de aceptación social.

integral? La cláusula de excepción nos indica que la teoría de la elección racional permite cierta tolerancia explicativa. Pero ¿qué tipo de tolerancia puede resistir? Al responder a esta pregunta, intentaré ofrecer una interpretación de la teoría de la elección racional los suficientemente amplia como para resultar empíricamente plausible, pero no tanto que se convierta en empíricamente carente de contenido (Pettit, 1993a). Formulo esta interpretación absolutamente bajo mi responsabilidad. Debo agregar que no se ofrece como una paráfrasis de lo que podrían haber sido las ideas de Harsanyi.

Existen dos tipos de casos con respecto a los cuales es necesario considerar esta cuestión. El primero es una situación en la cual los individuos debaten y deliberan, a menudo abiertamente, acerca de sus opciones –administran su conducta abiertamente– en términos más o menos interesados. Los contextos asimilables a mercados son ejemplos de ello. En tales contextos, la expectativa normal es que un individuo se vea atraído por determinada opción únicamente en la medida en que ésta lo beneficie más que otras alternativas, habitualmente en términos de beneficio económico. Podríamos decir que, en tales contextos, el discurso es predominantemente de regateo. Si una parte recomienda a otra una opción determinada, es siempre sobre la base de que lo beneficia más, en función de su interés egoísta, que cualquier alternativa posible.

¿Qué significa afirmar que la conducta humana, en tales situaciones asimilables a mercados, puede explicarse en gran medida en función del interés egoísta? Lo más plausible es que indique que las consideraciones que motivan deliberativamente a las personas a realizar una acción, en tales situaciones, se basan predominantemente en su interés egoísta. Puede ser que cada individuo esté sujeto, en cierta medida, a otras consideraciones distintas y que algunos individuos otorguen un lugar de relativa relevancia a estas razones no egocéntricas dentro de sus deliberaciones. Pero la idea es que las consideraciones egocéntricas desempeñan un papel decisivo en la determinación de lo que hacen los individuos, en su conjunto. Así, la teoría predice que cualquier cambio en las recompensas disponibles para los agentes en función del interés egoísta se traducirá en un cambio en la conducta agregada. La teoría puede explicar cualquiera de estas variaciones comparativas/estáticas.

Sin embargo, la cuestión de qué zonas grises deja sin explicar la teoría de la elección racional resulta mucho más complicada en las situaciones cuyo carácter no es asimilable a mercados. En la mayoría de los contextos sociales, el discurso en cuyos términos delibera y debate sus opciones la mayoría de las personas no es egocéntrico. Los individuos toman sus de-

cisiones y administran su conducta, no por referencia a su propio bienestar –o, al menos, no exclusivamente por ello– sino también al bienestar de su familia, de sus amigos o de una determinada entidad a la que pertenecen. Además, en ocasiones toman sus decisiones y administran su conducta sin tener en cuenta consideración de bienestar alguna. Piensan y deciden qué harán a la luz de consideraciones relacionadas con lo justo o equitativo, lo estéticamente agradable o divertido, lo que dejará una huella o mejorará su comprensión, lo que resultará adecuado al contexto, y así sucesivamente. Las posibilidades son infinitas.

¿Cómo puede la teoría de la elección racional, que invoca solamente consideraciones de interés egoísta, aspirar a explicar aquella conducta que es administrada y generada –exclusivamente, podríamos suponer– a la luz de formas no egocéntricas de reflexión y justificación, es decir, en términos de discursos que nos orientan hacia los intereses de otros o de discursos relativamente desinteresados? (Hindess, 1988). ¿En qué medida podemos esperar explicar esta conducta, mucho menos de manera completa? En otras palabras, ¿cómo podemos encontrar una interpretación adecuada aquí para la salvedad de que la teoría explica «en gran medida», o al menos parcialmente, tal conducta?

Una manera de responder a la dificultad sería afirmar que las personas nunca administran deliberativamente sus decisiones en términos no egocéntricos, exclusivamente o no. Pero esto no resulta convincente. Fuera del mercado, raramente se considera aceptable que los agentes tomen decisiones sobre la base de consideraciones de interés egoísta únicamente. El amigo, consejero o político que defiende determinadas iniciativas sobre la base de que resultan personalmente ventajosas pierde todo derecho al afecto, la atención o el respeto. Además, en áreas donde resulta socialmente inaceptable tomar decisiones sobre la base de consideraciones egoístas, no es probable que los individuos realicen sus elecciones sobre tales bases de todas maneras. Por supuesto, todos reconocemos que las personas, en ocasiones, pueden basar sus acciones en consideraciones que no resultan socialmente aceptables. Por ejemplo, los políticos pueden actuar con el propósito de lograr su reelección incluso cuando invocan fundamentos más altruistas para sus políticas. No obstante, a menos que adoptemos una visión totalmente desesperanzada de los seres humanos, debemos pensar que, en muchos casos, las personas realmente se guían por las justificaciones socialmente aceptables que invocan para sus acciones.

Una segunda respuesta, que es más generalmente aceptada pero no resulta mucho más atractiva, consistiría en afirmar que, aunque fuera del mercado los individuos no deliberan explícitamente en términos egoístas,

lo hacen implícita o inconscientemente, y que este hecho permite explicar su conducta en función de su interés. Gary Becker (1976, p. 7) sugiere que se inclina por este punto de vista: «el enfoque económico no supone que las unidades de decisión sean necesariamente conscientes de sus propios esfuerzos para maximizar [su bienestar] ni que puedan verbalizar o describir de alguna manera informativa las razones subyacentes a los patrones sistemáticos de su conducta. Esto resulta coherente con el énfasis en el subconsciente que encontramos en la psicología moderna». (Véase también McCullagh, 1991.)

Esta respuesta puede no estar tan reñida con las apariencias como la perspectiva calculadora, ya que no afirma que nos guiemos únicamente por consideraciones egoístas en la administración de nuestra conducta. No obstante, nos exige que aceptemos un relato muy controvertido acerca de qué nos motiva realmente en nuestra deliberación y qué afecta la administración de nuestra conducta, el cual se contradice con nuestro sentido inmediato de nosotros mismos y de los demás. Indudablemente, el relato se aplica a ciertas ocasiones e, indudablemente, aun los mejores de nosotros estamos sujetos ocasionalmente al autoengaño acerca de nuestras motivaciones. Pero la idea de que este relato se aplique a la mayoría de nosotros la mayoría del tiempo es extravagantemente poco plausible. Con este enfoque, el costo epistémico de aceptar la teoría de la elección racional sería, seguramente, demasiado alto, ya que nos exigiría una revisión demasiado profunda de nuestra concepción espontánea de la mayoría de los seres humanos.

Me gustaría ofrecer una tercera respuesta, más plausible, a la dificultad planteada, que es una manera de comprender la naturaleza condicionada de la teoría de la elección racional en su aplicación a los contextos pertinentes (Pettit, 1993a, cap. 5; 1993b). Esta respuesta niega que sea necesario que los agentes deliberen explícita o implícitamente en función de interés egoísta para que la teoría de la elección racional resulte aplicable a su conducta. Sostiene que la teoría de la elección racional resulta relevante en la medida en que las consideraciones del interés egoísta estén presentes en las deliberaciones de los agentes y en sus prácticas de administración virtualmente y no realmente.

Las consideraciones del interés egoísta se encuentran virtualmente presentes en las deliberaciones de los agentes si se dan las siguientes condiciones:

1. El agente hace lo que hace por determinadas razones no egocéntricas, de manera que el interés egoísta no tiene una presencia real, explícita ni implícita, en sus deliberaciones.

2. Lo que el agente hace es más o menos satisfactorio (el criterio de satisfacción puede ser una variable) en términos egoístas: contribuye razonablemente bien al interés egoísta.
3. Además, si lo que el agente hace en función de consideraciones no egocéntricas no resultase satisfactorio en este sentido, esto causaría que comenzara a pensar en función del interés egoísta y, con toda probabilidad, que ajustara su conducta en consecuencia.

¿Cómo podría ser verdadera la tercer proposición? ¿Cómo podría comprobarse que, en el caso de que la conducta no egocéntricamente justificada de un agente no sirviera a sus intereses egoístas satisfactoriamente, éste comenzara a replantearse lo que está haciendo? Tendría que darse el caso de que, a medida que la conducta se volviera egocéntricamente insatisfactoria, este hecho se registrara en la conciencia del agente y encendiera una luz de alerta. Supongamos que la conducta es egocéntricamente satisfactoria sólo en la medida que permita que el agente mantenga, sin especial esfuerzo, el estilo de vida que tienen aquéllos situados dentro de su grupo de referencia (Runciman, 1972). De ser así, el hecho de que el agente estuviera en aparente desventaja con respecto a sus pares en ciertos aspectos o el hecho de que fuera evidentemente necesario un gran esfuerzo a fin de mantener su posición encenderían una luz de alarma y harían que el agente se replanteara su conducta deliberativamente generada por consideraciones no egocéntricas; que se replanteara las donaciones que hace a la caridad, el pago escrupuloso de los impuestos, su generosidad hacia sus familiares o lo que fuera.[8]

La mejor interpretación de la teoría de la elección racional para contextos asimilables a mercados –la interpretación que se mantiene plausible sin volverse vacía– supone que la teoría atribuye al interés egoísta una influencia parcial en la generación de la conducta. La mejor interpretación de la teoría para los contextos distintos a mercados, como sugiero en estas páginas, supone que la teoría adscribe al interés egoísta una influencia virtual en la conformación de la conducta, es decir, en las deliberaciones de muchos, si no de todos los seres humanos. Esta interpretación requiere el aval de una perspectiva como el enfoque del grupo de referencia, ya que

8. Nótese que ser egoísta virtualmente –en este sentido– resulta compatible con ser también, por ejemplo, moralista virtualmente (es decir, tener una manera de ser que asegure que, en caso de que la deliberación del agente conduzca a ciertas formas de conducta inmoral, se enciendan otras luces de alarma y que el agente se replantee su conducta). Sólo surgirán problemas con la concesión prevista si existen situaciones en las cuales los agentes no puedan honrar simultáneamente ambos tipos de restricciones.

es necesario un relato que explique cuándo un patrón de conducta resulta egocéntricamente satisfactorio, es decir, no enciende la luz de alarma. No continuaré refiriéndome a la cuestión en estas páginas, suponiendo solamente que existe un contenido para el concepto de lo egocéntricamente satisfactorio.

Sin embargo, ¿resulta suficiente esta influencia virtual para la relevancia explicativa? El hecho de que el interés egoísta influya virtualmente –es decir, que tenga una presencia virtual en las deliberaciones de los individuos– significa que se trata solamente de una causa latente y no efectiva. Significa que espera, listo para desempeñar un papel causal si se enciende una luz de alerta, pero que en realidad no tiene efecto causal alguno. ¿Cómo es posible que el interés egoísta tenga relevancia explicativa para la conducta externa al mercado, si en realidad no cumple ningún papel en su producción?

Consideremos un ejemplo pertinente. Tomemos la explicación de la elección racional, hoy bien conocida, con respecto a por qué la esclavitud continuó firmemente vigente en el sur de Estados Unidos hasta la Guerra Civil: que consistía en un ordenamiento económico que recompensaba adecuadamente a los dueños de las plantaciones. Esta explicación se sugiere, por ejemplo, en el clásico texto de Fogel y Engerman (1974, p. 4): «La esclavitud no era un sistema que los dueños de las plantaciones mantenían irracionalmente porque no percibían o eran indiferentes a sus intereses económicos. La compra de un esclavo resultaba generalmente una inversión altamente rentable y generaba tasas de retorno que podían compararse favorablemente con las oportunidades de inversión más destacadas dentro de la industria». Supongamos que los dueños de las plantaciones no pensaran realmente demasiado en términos económicos acerca de sus compromisos. Supongamos que persistían en su conducta simplemente por hábito o por concebir tales compromisos, como ha sido sugerido, en términos morales, cuasi religiosos. Supongamos, en otras palabras, que el interés económico tenía como máximo una presencia virtual en sus deliberaciones. ¿Es posible invocar tal interés egoísta, de todos modos, para la explicación de su conducta?

No es posible invocarlo para explicar por qué surgió la conducta o por qué se reprodujo si suponemos que en realidad estaba motivada por el hábito o por consideraciones no egocéntricas. No obstante, podemos incorporar al interés egoísta en otra función explicativa, también importante. Podemos invocarlo para explicar por qué la conducta y el sistema que ésta generaba continuaron resistiendo, por qué tenía tal robustez que es posible afirmar que, incluso aunque los dueños de las plantaciones hubie-

ran comenzado a replantearse o revisar lo que hacían –como, por supuesto, deben haberlo hecho algunos–, es probable que aun así la conducta y el sistema permaneciesen vigentes. La idea es que, dada la manera en que satisfacía los intereses económicos egoístas, cualquier dueño de plantación que hubiese comenzado a cambiar su conducta se hubiese enfrentado rápidamente con un serio declive de su fortuna y, frente a tal perspectiva, hubiese estado inclinado a regresar al *statu quo*. Supongamos, para asumir la posibilidad contraria, que la tenencia de esclavos no hubiese satisfecho el interés económico egoísta de los terratenientes. La implicación del modelo es que el hecho de que a los propietarios individuales que repudiaran la esclavitud –al azar o como experimento– comenzara a irles mejor económicamente que a sus colegas anticuados, probablemente hubiese originado un abandono generalizado del sistema.

Con esto concluyo la referencia a la teoría de la elección racional. La teoría es un esquema heurístico de explicación que sugiere que el interés egoísta, económico o social, desempeña un importante papel en la producción de conductas humanas. A menudo, los individuos se basan en consideraciones del interés egoísta para su deliberación, como en el caso de los contextos de mercado. Cuando es así, tales consideraciones tienen una influencia por lo menos parcial –aunque indudablemente importante– en su conducta. Cuando los individuos no se basan en tales consideraciones, cuando deciden qué hacer sobre la base de otros tipos de deliberación, aun así el interés egoísta tiene una influencia virtual en sus decisiones. Esto significa, en general, que su conducta resultará por lo menos satisfactoria en función de su interés. Su conducta no ignorará al interés egoísta al punto de encender una luz de alerta.

2.3. Diseño institucional racional

Suponer que el diseño institucional desempeña un papel en la vida de los seres humanos, equivale a suponer que los individuos en general no se encuentran inevitablemente motivados a cumplir con las normas de conducta pertinentes en ausencia de posibles iniciativas de filtro o de sanción. Si así fuese, no tendría sentido intentar alterar las variables institucionales. De hecho, sería positivamente arriesgado aventurarse a cualquier intervención de este tipo, ya que interferir en una institución podría tener un efecto negativo sobre actuales niveles de desempeño que resultan satisfactorios.

La teoría de la elección racional tiene una explicación preparada para explicar por qué el cumplimiento institucional, como podríamos denominarlo, no resulta inevitable a partir de la espontaneidad. La explicación preferida no indica que las personas sean torpes y no se den cuenta de que el cumplimiento respondería al bien general, como podemos suponer. Tampoco sugiere que las personas sean propensas a tales excesos de emoción o pasión que se desvíen de las normas institucionales de una manera más o menos espasmódica. La explicación que ofrece la teoría de la elección racional –al menos como un relato parcial– consiste en que el interés egoísta a menudo aconseja una conducta de incumplimiento. La conducta exigida por el bien general no siempre es el tipo de conducta que promueve el interés egoísta individual. Por el contrario, la primera exige, en ocasiones, un grado de sacrificio.

Para explicar esta idea, la teoría de la elección racional pueden recurrir a los postulados de la teoría de los juegos sobre el Dilema del Prisionero. En este caso, las recompensas están estructuradas de manera tal que lo mejor para cada uno de los dos prisioneros es confesar un crimen cometido en común, independientemente de que el otro confiese o que se niegue a hacerlo. La confesión es el resultado de equilibrio y ninguno de los dos puede obtener un beneficio para sí mismo por desviarse de él. Además, es un resultado del cual ninguno de los dos puede desviarse sin una pérdida. A pesar de esto, resulta mejor para cada uno de los prisioneros que ambos se nieguen a confesar: la confesión conjunta es inferior, en el sentido de Pareto, a la negativa conjunta a confesar. Así podemos ver que una forma de conducta –la negativa conjunta a confesar– puede satisfacer el bien general sin constituir una motivación individual para los individuos egoístas involucrados.[9]

Dada esta explicación del incumplimiento, ¿qué sugiere la teoría de la elección racional como solución? ¿Qué sugiere con respecto al diseño institucional? Existen dos estrategias generales que podrían guiarnos en la investigación, dependiendo de si nos concentramos principalmente en los no cumplidores o en el hecho de que, aunque se produzca el incumplimiento, existen también muchos que cumplen con toda norma pertinente o que, al menos, están dispuestos a cumplirlas. El primer tipo de estrategia puede describirse como centrado en la desviación y el segundo, como centrado en el cumplimiento. Me propongo sostener que, aunque

9. Cabe señalar que incluso los altruistas pueden verse involucrados en un dilema del prisionero (véase Parfit, 1984; Pettit, 1985). Los individuos movidos por el interés egoísta son personas cuya motivación es relativa al agente: cada uno busca un bien que se define por referencia a quiénes son. Los individuos altruistas –en particular, los perfectamente altruistas– pueden también estar relativamente motivados con respecto a un agente con el efecto de encontrarse en el dilema del prisionero, ya que cada uno busca exclusivamente el bien de otro.

la primera estrategia sea la que más se destaca, la teoría de la elección racional debería inclinarse por la segunda.

2.3.1. La estrategia centrada en la desviación

Esta estrategia parte de la idea de que, dado que el interés egoísta desvía a los individuos –al menos, a algunos de ellos– de la conducta de cumplimiento, es necesario realizar intervenciones institucionales que aseguren que para tales individuos el cumplimiento se convierta, por el contrario, en la opción dictada por su interés egoísta. Debemos aumentar su motivación para cumplir, mejorando las recompensas que inducen al cumplimiento. Si la tasa de retorno esperada del interés egoísta es *X* para la desviación y algo menor para el cumplimiento, debemos introducir sanciones que aseguren que se restablezca el equilibrio, al menos en cierta medida. El diseño institucional debería guiarse por el objetivo de establecer motivadores que logren que un número creciente de individuos potencialmente incumplidores se mantenga en la senda deseada.

La manera ideal de implementar la estrategia centrada en la desviación sería identificar el motivador que se necesita para cada individuo –suponiendo que exista– y asegurarse de que esté en vigor. Por supuesto, este enfoque personalizado no resultará viable en nuestro mundo, ya que no es posible tener sanciones diferentes para los diferentes individuos. ¿Cómo podemos proceder con la estrategia, entonces? La respuesta obvia es que deberíamos considerar que el individuo es perfectamente egoísta y establecer sanciones que aseguren, como mínimo, que si tal individuo es culpable de una desviación, la sanción será suficiente para que se arrepienta de haberla cometido. Afirmo que «como mínimo» debemos asegurarnos esto porque el objetivo de desalentarlos –por medio de la incertidumbre sobre si los incumplidores serán atrapados y sentenciados– exigiría sanciones aún más severas.

La idea general para la estrategia centrada en la desviación, entonces, será brindar una motivación mayor de la que se necesita para la mayoría de los individuos –ciertamente, superior a la que bastaría para causar el arrepentimiento de alguien condenado– a fin de asegurar que la motivación resulte suficiente para todos. La idea vincula este enfoque con la estrategia orientada a los canallas que defienden autores como Hume y Mandeville. En las palabras del primero (Hume, 1875, pp. 117-118), al «establecer los diversos balances y contrapesos de la Constitución, cada hombre debería ser considerado como un canalla que no tiene otro fin para sus acciones

que su interés privado». O como ya había afirmado Mandeville (1731, p. 332), el mejor tipo de Constitución es aquel que «permanece incólume aunque la mayoría de los hombres demuestren ser canallas». La estrategia centrada en la desviación se reduce, en la práctica, a lo que en ocasiones se conoce como la estrategia para canallas.

Sin embargo, esta estrategia está sujeta a dos dificultades principales, especialmente dentro de la perspectiva de la teoría de la elección racional desarrollada en la ultima sección. (Véase Ayres y Braithwaite, 1992; Brennan y Buchanan, 1981; Goodin, 1992.) Una de ellas es la dificultad genérica de que, si deseamos establecer las sanciones o recompensas extremas que pueden ser necesarias para motivar a los canallas –es decir, si aplicamos sanciones centradas en la desviación– necesitaríamos basarnos en un sistema disciplinario centralizado que otorgue un gran poder a los responsables de la administración central de las sanciones. No obstante, es probable que en ese caso se creen más problemas de los que se resuelvan, ya que nos enfrentaríamos de una manera particularmente dramática con un desafío inmemorial: «¿*Quis custodiet costodes*?»; ¿Quién nos guarda de los guardianes? En especial, ¿quién nos protegerá de guardianes que han recibido un poder tan grande para sancionar y persuadir?

Aunque supongamos que sea posible evitar este problema, se presenta una segunda dificultad, más específica. Está implícita en tres proposiciones plausibles, todas las cuales constituyen predicciones de la perspectiva presentada en la última sección; la perspectiva virtual de la elección racional.

1. Aun de no existir sanciones extremas, centradas en la desviación, muchos agentes respetarán los patrones de conducta pertinentes sobre la base de un régimen de deliberación no egocéntrico. El cumplimiento les permite quedar en una posición lo suficientemente buena como para no encender una luz de alerta ni, por lo tanto, activar una reconsideración egocéntrica respecto de su conducta.
2. Un régimen egocéntrico de deliberación tendría menos probabilidades de generar el mismo nivel de cumplimiento entre los agentes pertinentes, incluso en presencia de sanciones centradas en la desviación.
3. La introducción de sanciones centradas en la desviación tendería a hacer que los agentes cambiaran su modalidad de deliberación no egocéntrica por una egocéntrica.

Conclusión: la introducción de sanciones centradas en la desviación tiene probabilidades de hacer más mal que bien.

La primera proposición nos señala que en muchas áreas pertinentes, tal como se argumentó en la última sección, las personas guían su conducta por consideraciones no egocéntricas que ofrecen razones categóricas para cumplir con los patrones en cuestión. «¿Por qué renunciar a las vacaciones para ayudar a estas personas? Son mis padres». «¿Por qué pasar tanto tiempo corrigiendo estos exámenes? Tengo que ser justo con los estudiantes». «¿Por qué ir a esta reunión tan aburrida? Es lo que se espera de los miembros». «¿Por qué no robar el reloj? No soy un criminal». No es mi intención sugerir que tales consideraciones resultan siempre convincentes, ni tampoco que pueden resultar efectivas de no existir sanciones que las respalden; profundizaré este tema más adelante. Lo que afirmo es que, a menudo, constituye el único tipo de consideraciones que las personas reconocen y que pueden servir para mantenerlas más o menos automáticamente en la senda a la que se dirigen. Una lección que nos ha enseñado la sociología –y sí, efectivamente, hacía falta que la aprendiéramos– es que a menudo actuamos bajo el control de pilotos no egocéntricos, relacionados con los papeles sociales. Con frecuencia nos amoldamos al perfil del *Homo sociologicus*. Para tomar un ejemplo pertinente: una idea ampliamente aceptada señala que, en la medida en que las personas evitan el crimen, lo hacen debido a que las consideraciones por las cuales se guían convierten el crimen en algo impensable, lo eliminan de la lista de las alternativas pertinentes (Braithwaite, 1989).[10]

La segunda proposición se refiere al efecto probable de un aumento de la deliberación egocéntrica sobre el cumplimiento, es decir, a la administración egocéntrica de la conducta. El supuesto señala que, aunque existan sólidas sanciones egocéntricas que favorezcan el cumplimiento, de todas maneras un régimen de deliberación egocéntrica genera niveles de cumplimiento menores. En este caso, existen dos consideraciones particularmente reveladoras. La primera es el hecho de que, en el mejor de los casos, las consideraciones egocéntricas avalan el cumplimiento sólo de manera condicional y no categóricamente. Dentro de una modalidad de razonamiento no egocéntrica, que dependa de los papeles sociales, el cumplimiento se sostiene más o menos automáticamente, como hemos descrito anteriormente, y la cuestión de si vale la pena no se plantea siquiera (Durkheim, 1961). No obstante, bajo un régimen de deliberación egocéntrica, la opción entre cumplir o no hacerlo generalmente se hará sentir. Aunque la pregunta tenga a menudo una respuesta positiva, el hecho mismo de que se plantee en cada caso convierte al incumplimiento en una posibilidad más conspicua y probable.

La restante consideración señala que la deliberación egocéntrica sostiene solamente de manera condicional el cumplimiento y, además, una de sus condiciones es que se considere razonablemente alta la posibilidad de ser detectado. Esta es una debilidad particular, ya que existen tantos casos, en todos los ámbitos de la vida, en los que resulta posible que el contraventor o el incumplidor evite ser detectado y en los que basta un mínimo de reflexión –centrada en el interés egocéntrico– para darse cuenta de ello. Si un patrón de deliberación egoísta llevara a los individuos a evaluar, en cada caso, la probabilidad de que se detecte su incumplimiento, resultaría también probable que opten por desviarse. La práctica de tener en cuenta la probabilidad de ser detectado daría fácilmente lugar a hábitos de desviación, dado que ésta tiende a ser baja (Zimring y Hawkins, 1973).

La tercera proposición indica que, de implementarse la estrategia centrada en la desviación así como las penas más duras y las recompensas más altas que propone, probablemente esto causaría que muchos individuos con patrones de deliberación no egocéntrica los cambiaran por una modalidad egocéntrica. Esto se produciría, por ejemplo, al encenderse la luz de alerta sugiriendo a los individuos que sus patrones establecidos de conducta no los satisfacen tanto como sería posible. Supongamos que descubro que el sueldo por mi tipo de trabajo ha aumentado dramáticamente o que se endureció drásticamente la sanción por mentir en la declaración de impuestos que suelo presentar. Un efecto posible es que comience a preguntarme si no he estado cotizándome demasiado bajo, si no he estado dedicando un nivel excesivo de esfuerzo al trabajo o ignorando oportunidades de evadir impuestos que mis pares aprovechan regularmente. Al hacer que me plantee estas dudas, las nuevas sanciones pueden hacer que preste un grado inusitado de atención a la promoción de mi ventaja personal.

No obstante, existen otras maneras de comprobar la tercera proposición que son consistentes con la perspectiva del egocentrismo virtual. Una de ellas consiste en que el establecimiento de las nuevas sanciones –las grandes recompensas o las penas severas– puede hacer que las consideraciones egocéntricas se destaquen en mayor medida que con anterio-

10. Según la teoría de la elección racional, tal como se la interpreta en la última sección, este patrón resultará resistente en la medida en que sea lo suficientemente satisfactorio desde el punto de vista egocéntrico (o aparente serlo) para no encender una luz de alerta: es decir, sólo en la medida en que parezca satisfacer adecuadamente al interés egoísta. Vale la pena señalar que, pese a todo lo que hemos afirmado, el patrón puede en realidad satisfacer el interés egoísta mejor que la deliberación egocéntrica. Al igual que la honestidad puede ser la mejor política, un régimen de interés egoísta virtual puede resultar óptimo desde el punto de vista egocéntrico.

ridad, aunque no enciendan una luz de alerta, con lo que podrían eliminar o marginar los pensamientos no egocéntricos. Todas las sanciones se pagan en moneda egocéntrica, que representa recompensas o penalidades que responden al interés egoísta. El hecho de que se establezcan en una situación dada puede tener, en sí mismo, el efecto de inclinar las mentes de los individuos hacia una orientación egocéntrica. Las sanciones económicas o sociales que son lo suficientemente altas o severas como para motivar a un canalla pueden resultar tan altas o severas que eclipsen o socaven otras consideraciones en la deliberación de los agentes normales. Acostumbrados a concebir y tomar sus decisiones en términos más o menos profesionales o prudentes, por ejemplo, tales agentes pueden verse impulsados a pensar de una manera más interesada y orientada hacia los resultados como efecto de las nuevas sanciones (Lepper y Greene, 1978; Ayres y Braithwaite, 1992, pp. 49-51).

Otra posibilidad es que las sanciones centradas en la desviación no eliminen o desplacen a los pensamientos no egocéntricos, sino que los vuelvan menos convincentes. Es posible que tengan este efecto en la medida en que sirven a los agentes como un indicio de las actitudes de los demás, en particular las autoridades. El hecho de que se establezcan determinadas sanciones extremas para personas que se encuentran, generalmente, dentro de un área dada de conducta sugiere que los agentes pertinentes son tan egocéntricos en sus deliberaciones que no tendrían una conducta de cumplimiento si no existieran tales recompensas o penalidades. Y la proyección de tal expectativa puede convertirse en una profecía de autocumplimiento. Puede servir para legitimar la gestión egocéntrica de la conducta, al representarla como estadísticamente normal, lo cual puede hacer que los individuos se vuelvan más egocéntricos en sus hábitos. Este efecto se verá reforzado si las sanciones se toman como una señal de falta de confianza o estima por parte de los responsables. No es necesaria una excesiva imaginación para concebir una situación en la que alguien extremadamente profesional o puntilloso con respecto a sus niveles de desempeño (por ejemplo, en algo tan trivial como cumplir su horario) se vea llevado a pensar en términos del interés egoísta que proyecta la imposición de penas severas a determinadas infracciones (tales como llegar tarde). «Si creen que soy un egoísta, ya van a ver qué hace un egoísta» (Braithwaite y Makkai, 1994; véase también Williamson, 1983).

Hasta ahora hemos considerado tres casos en los cuales las sanciones centradas en la desviación pueden hacer que agentes no egocéntricos cambien por una modalidad egocéntrica de deliberación. Se puede encender

una luz de alerta, se pueden eliminar o marginar los pensamientos no egocéntricos, o se pueden transmitir actitudes desmoralizadoras de los demás, es decir, actitudes que socavan la deliberación no egocéntrica. Todos estos son efectos que hacen que las sanciones reduzcan el impacto de las consideraciones no egocéntricas en agentes bien predispuestos en los demás sentidos. Existen otros dos efectos del mismo tipo que deberíamos tener también en cuenta. Ambos actúan como una señal y son similares a la última posibilidad mencionada.

El cuarto efecto desvía la atención de los agentes hacia la accesibilidad de la opción de desertar –que anteriormente puede haberles parecido sólo una posibilidad marginal– o hacia determinadas maneras específicas de desertar. Consideremos el caso de un trabajador a quien se obliga a vigilar constantemente su reloj a través de la imposición de sanciones severas por llegar tarde a la oficina. Al obsesionarse con el reloj, perderá no sólo el compromiso personal que sentía antes de la desmoralización, sino que también comenzará a ser consciente de posibles pretextos para la satisfacción de su interés egoísta, que hasta entonces había ignorado. Es posible que comience a descubrir formas cada vez menos fatigosas de cumplir con lo que se le exige, o de aparentar cumplirlo.

El quinto y último efecto que deseo mencionar es el efecto de señal, por el cual los agentes cumplidores descubren, cuando se establecen sanciones centradas en la desviación, que hay otros que no lo son o que, al menos, no siempre han cumplido en la misma medida. Esta información socavará por sí misma el cumplimiento, en la medida en que éste incluye un elemento contractual tácito: cada uno hace su pequeña contribución en el entendimiento de que los demás hacen su parte también (Levi, 1987). Si la implantación de mayores recompensas o penas más severas constituye una señal para quienes son cumplidores de que hasta ese momento los otros se han aprovechado de ellos –de que han estado haciendo más que lo que les correspondía– esto puede provocar que relajen los esfuerzos que han venido haciendo, a pesar de las mayores sanciones que se hayan puesto en vigor.

Además de estos cinco efectos que reducen el impacto de las consideraciones no egocéntricas sobre los cumplidores, la implantación de sanciones centradas en la desviación puede tener también ciertos efectos adversos en la selección. Las penas más severas pueden disuadir a los cumplidores espontáneos de la idea de ingresar en determinado campo de actividad (es decir, pueden reducir el perfil idealista de esa ocupación), mientras que las recompensas más altas pueden atraer a agentes de mentalidad egocéntrica que previamente podrían haberla considerado un terreno inadecuado pa-

ra ellos.[11] Para ilustrar el efecto de tales recompensas, supongamos que existen personas que están más inclinadas que otras a comprometerse de una manera completa y consciente con el papel de médico, de investigador, de administrador o de político. Manteniéndose todo lo demás constante, podemos esperar que tales personas se sientan más atraídas hacia los puestos relacionados que los individuos con una disposición menos adecuada. No obstante, si las recompensas asociadas a tales puestos se hacen relativamente altas, todo lo demás deja de ser constante y es muy posible que descubramos que aquellos que son atraídos por tales puestos –y quienes los obtienen– incluyan una proporción creciente de personas que no están particularmente predispuestas a interiorizar los papeles sociales pertinentes. Es posible que descubramos que comienza a cubrirse los puestos con un número cada vez mayor de individuos interesados sólo en el dinero y los honores. Esta idea evoca la proposición que enfatiza Richard Titmuss (1971) en su defensa de la distinción entre la donación de sangre y su venta: es posible que quienes se sienten inclinados a donar sangre resulten una mejor opción como fuente de sangre saludable que aquellos tentados o forzados a venderla.

Espero que los diversos efectos tratados aquí se consideren relativamente plausibles. Es fácil ver cómo podrían materializarse en una variedad de áreas dentro de la vida social e institucional, con la imposición de penas centradas en la desviación. Podemos ver cómo una persona que respeta automáticamente el derecho penal podría dejar de considerarlo como algo con lo que se identifica y podría comenzar a buscar oportunidades estratégicas para burlarlo (Braithwaite y Pettit, 1992). Es posible ver cómo los políticos que perciben que su propia imagen es la de personas poco dignas de confianza –tan acorralados se ven por las reglamentaciones y las amenazas– podrían comenzar a responder a esa imagen, procurando encontrar ocasiones para su propio beneficio. Podemos ver cómo el investigador que se ve importunado y alienado por un comité de ética oficioso puede llegar a ser menos escrupuloso que hasta ese momento en el respeto de los principios éticos (Pettit, 1992). Y podemos ver cómo el gerente de una fábrica o de un restaurante podría adoptar una posición adversativa frente a los inspectores, a considerarse un adversario determinado a ganar algunas batallas, si el parte de la inspección resulta demasiado draconiano (véanse otros ejemplos en Bardach y Kagan, 1982; Ayres y Braithwaite, 1992).

11. En este punto, debo expresar mi gratitud por el diálogo con Geoffrey Brennan. Véase Brennan y Pettit, 1991.

La difusión de estos diversos efectos puede ilustrarse también por referencia a las recompensas elevadas, en lugar de las penas severas. He aquí un ejemplo proveniente del área de la investigación científica o académica. La inclinación natural de un investigador dedicado es invertir sus energías en problemas que le interesen o que parezcan intelectualmente prometedores. Las altas recompensas pueden actuar como interferencia y es posible que ceda a una conducta estratégica o especuladora con respecto a los proyectos que emprende. El tipo de mercado de inversiones que se asocia con la investigación tradicional puede ser reemplazado por un mercado de carácter especulativo. Un mercado en el cual cada uno avanza en la dirección que le resulta intelectualmente atractiva, con sólo una atención superficial a los precios –es decir, a recompensas que responden al interés egoísta–, puede ceder paso a un mercado en el cual cada uno intente moverse en la dirección que, dentro de la nueva percepción, logre el precio más alto posteriormente. Estos precios o recompensas de distracción pueden ser económicos, ya que cada investigador procura acceder al área que, por ejemplo, tiene más probabilidades de atraer la atención de los organismos que aportan fondos. O bien pueden ser sociales, a medida que cada uno procura aventajar al rebaño al apadrinar ideas que prometen ponerse de moda para lograr la atención y el aplauso públicos. La cultura intelectual de París, al menos como a menudo ha sido parodiada, constituiría un mercado especulativo de este último tipo, dominado por pronósticos sobre la dirección en la que se orientará el rebaño, a diferencia del mercado de inversión que representan los ámbitos académicos más tradicionales.

Con esto se cierra este tratamiento de estas tres proposiciones y la dificultad que implican para la estrategia centrada en la desviación dentro del diseño institucional. Comprendo que ninguna de las proposiciones ha sido demostrada en estas páginas: todas ellas constituyen supuestos empíricos más o menos vulnerables. No obstante, el hecho mismo de que la dificultad que nos señalan constituya una posibilidad real debería plantear dudas sobre la conveniencia de proceder con la estrategia centrada en la desviación y llevarnos a que nos preguntemos si existe alguna otra estrategia que pueda evitar esa dificultad. Me aboco ahora a la consideración de una alternativa que parecería lograrlo.

2.3.2. La estrategia centrada en el cumplimiento

La estrategia centrada en la desviación se guía por la necesidad de lidiar con el canalla: es decir, con la persona más explícitamente egoísta que

existe. La estrategia centrada en el cumplimiento se guía por la necesidad de tratar a una clase de individuos más habitual: alguien que delibera de una forma no egocéntrica en la mayoría de los contextos y que sólo se centra en su interés egoísta de una manera asociada con la presencia virtual de ese interés. La idea es que el diseño institucional, en primer lugar, debería obrar a partir de la disposición positiva de este tipo de personas y considerar cómo enfrentar a aquellos que son más explícitamente egoístas únicamente en segundo lugar. Debería sustentarse sobre su punto fuerte, buscando los medios para estabilizar las disposiciones que se inclinan por el cumplimiento, y sólo después ocuparse de cómo compensar las debilidades, cómo protegerse contra los problemas a los que puede dar lugar el régimen deliberativo del interés egoísta.

Kant es pesimista acerca de que fuera posible que surgiera algo derecho a partir de la retorcida madera de la que está hecha la humanidad. Aunque su pesimismo sea oportuno, debemos tener en claro que resulta más probable que nos aproximemos a la rectitud en el caso de algunas muestras de madera humana que en el de otras. La estrategia centrada en el cumplimiento se toma a pecho esta enseñanza y afirma que debemos concentrar nuestra atención en los mejores o más dúctiles ejemplos primero para luego ocuparnos de cómo encuadrar a aquellas piezas que son particularmente retorcidas. He descrito la primera estrategia como centrada en la desviación debido a que se guía por el supuesto de que el cumplimiento exige, principalmente, de recursos motivacionales extraordinarios para controlar a los que se desvían, a los «canallas». He descrito la segunda como centrada en el cumplimiento ya que la premisa, en este caso, indica que el primer requisito para el cumplimiento consiste en abstenerse de perturbar las prácticas de deliberación o administración que mantienen en la buena senda a los cumplidores.

Presentaré esta estrategia centrada en el cumplimiento a través de tres principios. El primero indica que deben explorarse las posibilidades de establecer mecanismos de filtro antes de considerar las opciones de sanción; el segundo, que los mecanismos de sanción establecidos deben ser, en la medida de lo posible, propicios a la deliberación no egocéntrica; el tercero, que los mecanismos de sanción deberían resultar también eficaces como motivación.

2.3.3. Primer principio: el filtro antes de la sanción

El primer principio establece que, en el diseño institucional, debemos concentrarnos en las posibilidades de establecer filtros antes de explorar las sanciones. El principio se ve avalado por nuestra reflexión acerca de los problemas a los que dan lugar las sanciones excesivas, ya sean penas o recompensas. Si es posible filtrar a la población de agentes pertinente para un ejemplo dado del diseño institucional de manera tal que, en general, aquellos involucrados no estén movidos deliberativamente por el interés egoísta, sino inclinados a deliberar acerca de sus opciones en la moneda que resulte contextualmente adecuada a la elección en cuestión, entonces resulta posible asegurar el grado deseado de cumplimiento sin recurrir a sanciones severas y peligrosas. Asimismo, si es posible eliminar las opciones perjudiciales relacionadas con un ejemplo de diseño institucional de la lista de alternativas disponibles o, lo que resultaría más atractivo, si pueden incorporarse a la lista las opciones adecuadas, entonces es posible inducir los individuos a que actúen adecuadamente sin la intervención de tales sanciones. Por supuesto, las oportunidades de filtrado no siempre estarán disponibles e, incluso cuando es así, pueden resultar demasiado costosas como para ser realmente viables. No obstante, de estar disponibles y de ser realmente viables, el primer principio determina que los diseñadores institucionales deben tratar de aprovecharlas.

El mecanismo de filtro más comúnmente reconocido es el que se centra en el agente y tiene el propósito de eliminar a determinados individuos de un marco dado. Un buen ejemplo es el proceso de eliminación de los potenciales miembros de un jurado, a fines de asegurar que no se admita a ningún amigo ni enemigo del acusado, ni nadie con un prejuicio en su favor o en su contra. Si estamos tratando con este tipo de grupo depurado de personas, podemos ser relativamente optimistas en cuanto a que se conformarán a la norma que determina que deben intentar decidir a conciencia si las pruebas demuestran la culpabilidad del acusado más allá de toda duda razonable. De no ser así, se presentaría todo tipo de peligros y podría creerse que sólo una forma draconiana de sanción –con todas las dificultades que ya hemos estudiado– puede ofrecer alguna esperanza de mantener a los jurados en la buena senda.

La manera en que se seleccione un determinado grupo será dictada, en buena medida, por el tipo de motores –incluyendo el tipo de sanciones– que esperamos que influyan sobre los agentes en cuestión. Supongamos que tenemos la esperanza de que el jurado generalmente se vea movido por el valor de la deliberación a conciencia, en el cual la voz virtual del interés egoísta aparece atenuada, y que, caso contrario, se los sancione con la desaprobación que probablemente sientan otros por un enfoque más indolente (Pettit, 1993a; Brennan y Pettit, 1993). En este caso, el filtro no sólo se establece para eliminar a cualquiera que tenga un interés personal en el

resultado. También procuraremos asegurarnos de que se seleccione un grupo heterogéneo de jurados –de manera tal que lo que atraiga la desaprobación sea la actitud realmente indolente y que la verdadera conciencia atraiga la aprobación– e intentaremos que el filtro admita a los jurados adecuados. Si el grupo de jurados compartiese antecedentes o circunstancias similares y si fuese previsible que un individuo de tal extracción llegase a un determinado juicio acerca del acusado, es posible que sea más fácil para un individuo ganar su aprobación si responde a tales expectativas que si actúa guiado por su propia conciencia.

La segunda posibilidad de selección consiste en trabajar sobre las opciones disponibles en lugar de centrarse en los agentes. Este tipo de mecanismo de filtro puede funcionar a través de un organismo de financiamiento o de admisión; por la acción de normas o criterios de elegibilidad; o bien en una serie de maneras menos obvias. Los comités de ética cumplen un papel selectivo con relación a ciertos tipos de investigación. La norma que establece que nadie puede ser presidente de Estados Unidos por tres períodos consecutivos desempeña un papel similar en relación con determinados proyectos políticos. Existen otros mecanismos que desempeñan el mismo tipo de función en las más diversas áreas. Consideremos, por ejemplo, el mecanismo de balance de poderes, que exige que cualquier instrumento legislativo sea aprobado por dos o más cuerpos independientemente; dos o más cuerpos que representan intereses relativamente opuestos. Esto asegura que exista un mecanismo de filtro que elimine cualquier opción que dañe u ofenda profundamente a cualquiera de esos intereses.

Todos estos son ejemplos de mecanismos de filtro para eliminar ciertas opciones. Pero los mecanismos que admiten determinadas opciones también representan una posibilidad prominente en el diseño institucional. Consideremos las situaciones en las cuales los individuos reciben recursos para que denuncien a determinadas autoridades o, más genéricamente, para que presenten demandas y apelaciones. Representan mecanismos para que las personas comunes seleccionen opciones que puedan funcionar como importantes controles sobre la conducta de las autoridades (Peters y Branch, 1972; McCubbins y Schwartz, 1984). El principio subyacente nos señala una gama de otros ejemplos posibles con el mismo efecto. Consideremos todos los ordenamientos, por ejemplo, que protegen a las personas comunes contra la interferencia o la explotación al asegurarse que continúen teniendo acceso a determinadas opciones: la opción de ser representadas por un abogado, a través de la asistencia jurídica estatal; la opción de ser hospitalizadas, a través de Medicare (el sistema de Salud Pública gratuito de Estados Unidos); la opción de conocer qué factores influyen en determinadas decisiones oficiales, a través de las leyes de libertad de información; y así sucesivamente.

Al igual que en el caso de los mecanismos de filtro centrados en los agentes, el filtrado de opciones nos permitiría evitar el recurso a sanciones excesivas. Dada la línea argumental que hemos seguido en contra de una estrategia centrada en la desviación, la importancia de explorar con detenimiento todas las medidas de filtro posibles –antes de recurrir a las sanciones que sean necesarias– parecería una cuestión del más elemental sentido común.

James Madison expresa el espíritu de nuestro primer principio cuando escribe, en el documento n.° 57 de *El Federalista*: «el objetivo de toda Constitución política es, o debería ser, en primer lugar, ganar como gobernantes a hombres que posean una mayor sabiduría para discernir y una mayor virtud para buscar el bien común de la sociedad; en segundo término, que se tomen las precauciones más efectivas para mantenerlos virtuosos, mientras continúen siendo depositarios de la confianza pública» (Wills, 1982, p. 289). El primer principio implementa lo que para Morton White es la idea guía de Madison en el diseño institucional: que deberíamos tomar las diferentes motivaciones de los diferentes individuos y grupos como dadas y, luego, tratar de que las oportunidades se correspondan con las motivaciones de manera que promuevan mejor el bien común (White, 1987).

2.3.4. Segundo principio: sancionar, pero de una manera deliberativamente alentadora

Ya nos hemos referido a la proposición que afirma que, en el diseño institucional, debemos concentrarnos en las iniciativas de filtro antes de efectuar cualquier intervención de sanción. El segundo principio que asocio con la estrategia centrada en el cumplimiento dispone que debemos buscar mecanismos de sanción además de los de filtros pero, en particular, mecanismos de sanción que resulten deliberativamente alentadores. Las sanciones deliberativamente alentadoras, en cualquier área, son aquellas que tienden a reforzar el tipo de hábitos deliberativos que constituyen o producen la conducta deseada.

Dada la importancia de los filtros, ¿nos permiten evitar completamente el empleo de sanciones en nuestro diseño de instituciones? Podrían hacerlo en el improbable caso de que eliminaran todas las opciones perjudi-

ciales de la lista de alternativas disponibles o que excluyeran a todos los individuos peligrosos del conjunto de los agentes. Sin embargo, ¿podrían hacerlo en situaciones más habituales, en las que siempre existe la posibilidad del incumplimiento? No lo creo así. Dos consideraciones nos demuestran, desde un punto de vista interno dentro de la teoría de la elección racional, que continuará siendo siempre necesario basarse en mecanismos de sanción además de los de filtro.

La primera consideración es la siguiente. Aunque determinados agentes estén dispuestos a deliberar de una manera que genere la conducta deseada de manera confiable, en cierto contexto, la ausencia de sanciones que castiguen la conducta contraria (de cualquier sanción basada en intereses particulares) puede desviarlos de la buena senda. Consideremos el relato del anillo de Gyges, en el cual se nos pide que nos preguntemos si continuaríamos comprometidos con la virtud aunque poseyéramos un anillo que nos hiciera invisibles y que nos permitiera recurrir impunemente a conductas más viles. La ausencia de sanciones que propone este relato es lo que hace tan plausible que incluso un agente muy virtuoso pueda ser corrompido. La ausencia de sanciones en esta disposición, exclusivamente de filtro, que se nos pide que imaginemos debería asimismo obligarnos a reflexionar antes de avalar la propuesta en cuestión. Nada obsta a que sean verdaderas, simultáneamente, la idea de que un agente ignora las sanciones existentes cuando busca razones para adoptar una determinada forma de conducta deseada y la idea de que la ausencia de tales sanciones causaría que se desviase de ella. Esto resultaría evidente dentro del modelo virtual de interés egoísta propuesto. La ausencia de toda sanción haría evidente la existencia de una forma alternativa de conducta que satisface mejor al interés egoísta, es decir, que encendería una luz de alerta; al volverse conspicua esta opción, resulta completamente posible que el agente se vea atraído hacia el incumplimiento (Braithwaite y Pettit, 1990, pp. 139-140).

Existe una segunda razón para recurrir a las sanciones además de a los filtros, que también exige atención dentro del tipo de perspectiva de la elección racional que hemos adoptado. Supongamos que un determinado agente se ve llevado deliberativamente a adoptar una determinada forma de conducta deseada en un contexto en el cual se han eliminado todas las sanciones que la refuerzan. Aunque la ausencia de sanciones no lleve al agente a considerar si debería promover su propio interés egoísta a través de la desviación, puede llevarlo a preguntarse si los demás continuarán cumpliendo su parte. La ausencia de sanciones puede significar que el agente pierda todo sentido de certeza al respecto y que la futilidad de hacer un esfuerzo aislado (un tipo de contribución que resulta muchas veces

inútil) lo lleve a desviarse de la senda trazada. La idea se ve reforzada por la memoria de Chester Bowles, un funcionario a cargo de la regulación del comercio de Estados Unidos durante la Segunda Guerra Mundial: el 20% de las empresas cumplía incondicionalmente cualquier norma, el 5% intentaba evadirla, y el 75% restante tendía a cumplirla, siempre y cuando el 5% que no cumplía fuera expuesto a una amenaza creíble de detección y castigo (Bardach y Kagan, 1982, pp. 65-66; véase también Levi, 1987).

Sin embargo, si bien son necesarios los mecanismos de sanción de cierto tipo, la segunda proposición insiste en que deberían ser deliberativamente alentadores en su carácter. He afirmado ya que las recompensas altas y las penas severas pueden activar la deliberación interesada por parte de agentes que, de otra manera, se guiarían por consideraciones no egocéntricas –consideraciones, podemos suponer, que alentarían el cumplimiento– y que, al hacerlo, pueden conducir a los agentes al incumplimiento. La lección que nos deja ese argumento es que el diseño institucional debería tratar de evitar las recompensas altas o las penas severas que tengan probabilidades de arrojar un efecto deliberativo disruptivo como ese. Debería inclinarse por las sanciones que tendieran a preservar e incluso reforzar la deliberación no egocéntrica.

Nuestro análisis de los casos en los que las sanciones centradas en la desviación pueden interferir con la deliberación no egocéntrica nos señala ciertas restricciones que deben satisfacer las sanciones. Ya sean recompensas o penas, las sanciones deberían ser cuantitativa y cualitativamente tales que eviten crear condiciones como las siguientes:

1. Las sanciones encienden una luz de alerta para los agentes, presentándoles su situación existente de cumplimiento como egocéntricamente insatisfactoria.
2. Las sanciones eliminan o vuelven marginales los pensamientos no egocéntricos; concentran la atención de los agentes en cuestiones más orientadas al interés egoísta.
3. Las sanciones convierten en menos atractivos los pensamientos no egocéntricos, al sugerir a los agentes que los demás, en especial otras personas que detentan autoridad, tienen un mal concepto de ellos o no confían en ellos.
4. Las sanciones dirigen la atención de los agentes hacia posibilidades de no cumplimiento que previamente podrían no habérseles ocurrido nunca.
5. Las sanciones sugieren a determinados agentes que otros no han estado cumpliendo con su parte, y que ellos están llevando la carga del cumplimiento sin el apoyo de los demás.

6. Las sanciones tienen efectos selectivos que desalientan a aquellos que tienden a ser naturalmente más cumplidores o que atraen hacia el área pertinente a aquellos que están menos naturalmente inclinados a cumplir.

La mejor manera de demostrar cómo las sanciones pueden ser deliberativamente alentadoras y cómo pueden evitar condiciones como las que se enumeran, sería recurrir a un ejemplo. Tomemos por caso cualquier comité en el cual el patrón de conducta deseado sea el voto a conciencia. Supongamos que hemos seleccionado a los miembros del organismo en cuestión a través de un mecanismo de recusación, de forma tal que, idealmente, nadie tenga un interés especial en el resultado y supongamos que nos hemos asegurado de que las labores del comité sean confidenciales, de manera que nadie se vea particularmente motivado por el temor a terceros con tales intereses. Estas medidas de filtro hacen probable que la mayoría de los miembros del comité se dedique espontáneamente a la cuestión que tienen por delante e intente tomar una decisión a conciencia; que, habiéndose sofocado la voz del interés egoísta, los individuos mostrarán el tipo de deliberación no egocéntrica contextualmente pertinente que asegura el voto a conciencia.

¿Qué tipo de sanciones alentaría este patrón de deliberación? Es costumbre en todo comité que el presidente invite a los miembros a exponer sus inclinaciones y a defenderlas ante los demás. Supongamos que los miembros, al defender su posición, deban hacerlo en términos cuyo interés no dependa de una perspectiva particular y sectorial, ya que el comité no está inclinado a favor de ninguna perspectiva de este tipo. A menos que los miembros puedan ofrecer una buena justificación de cómo se proponen votar –una justificación que sea convincente para las distintas perspectivas–, quedarán mal parados frente a los demás; parecerán necios o prejuiciosos. Así, podemos ver que en la situación descrita funcionan sanciones que tienen la capacidad de mantener en línea a cualquiera que se vea inclinado a desviarse: por ejemplo, cualquiera que esté impaciente por el tiempo que insume la reunión y que anuncie sus ideas de una manera apremiante (Brennan y Pettit, 1990; 1993).

Las sanciones basadas en la aprobación que aparecen aquí ofrecen un buen ejemplo de sanciones que alientan el tipo de deliberación que normalmente produce el resultado deseado. Si alguien espera quedar mal parado por no atenerse a un razonamiento serio o quedar bien parado por hacer esfuerzos eficaces en esa dirección, la observación no tendería a interferir en la deliberación y el discurso en cuestión. Por el contrario, si la

persona se ve motivada por tal sanción y es razonable en cuanto a la mejor manera de lograr la recompensa ofrecida y de evitar la pena, debería adoptar la conducta que le asegure esos resultados. No debería verse llevada, por ejemplo, a concentrarse explícitamente en la buena opinión de los demás, procurándola por todos los medios disponibles en un proceso de deliberación egocéntrica. El hacerlo la llevaría probablemente a ser descubierta, y la manera más segura de perder la consideración de los demás es quedar expuesto como alguien que se esfuerza por lograrla. «El axioma general en este terreno», como afirma Jon Elster (1983, p. 66), «es que nada resulta tan poco impresionante como la conducta diseñada para impresionar».[12]

Las sanciones que operan en el caso del comité son de un tipo que, en principio, puede movilizarse en cualquier foro en el que haya un debate por conducir y una decisión colectiva por tomar. Puede ser la sanción a la que se refiere Jürgen Habermas –quizás con demasiado optimismo– en su visión de los efectos de la situación de habla ideal. El autor imagina que, a medida que distintas partes van presentando razones a favor y en contra de diferentes opciones, se verán obligadas a argumentar en términos no sectoriales del tipo que puede interesar a cualquier persona, y que, al hacerlo, se inclinarán cada vez más a interiorizar el hábito con lo que se convertirán en contribuyentes verdaderamente imparciales y senatoriales (Elster, 1986b; Pettit, 1982; Goodin, 1992, cap. 7). Una razón por la que pueden verse obligados a argumentar en estos términos –incluso aunque no estuvieran espontáneamente inclinados a hacerlo– radica en que si no lo hacen no pueden gozar de la aceptación y aprobación de sus colegas dentro de ese foro.

Sin llegar a la sanción discursiva que contemplamos en estas páginas –es posible que este tipo de sanción no esté disponible a menudo– existen otras maneras de intentar asegurarnos de que las sanciones previstas en el diseño institucional alienten la deliberación deseada. Consideremos, por ejemplo, las sanciones que prevén los sistemas del derecho penal. Tradicionalmente, éstas han sido muy poco alentadoras de la deliberación más o menos moral que nos mantiene a la mayoría de nosotros del lado correcto de la ley, al menos en los países de Occidente (Braithwaite, 1989). No obstante, en principio no existe razón para que el derecho penal no comience a explorar posibilidades de sanción que contengan un aspecto

12. En este caso, entonces, la manera de maximizar el interés egoísta involucrado puede ser evitar la deliberación egocéntrica en términos de interés; es decir, mantener al interés egoísta como virtual. Véase nota al pie n.° 8.

de reprobación y que tiendan a alentar la deliberación deseada. Con John Braithwaite hemos argumentado en esta línea a favor del recurso a «la institución socializante, que busca hacer comprender a las personas la vergüenza del crimen induciéndolas a tener, no sólo las disposiciones conductistas, sino también los hábitos deliberativos de los ciudadanos virtuosos» (Braithwaite y Pettit, 1990, pp. 88-89).

2.3.5. Tercer principio: sanciones estructurales para lidiar con los ocasionales canallas

Hemos visto que, dentro de la estrategia del diseño institucional que se centra en el cumplimiento, la teoría de la elección racional recomendaría que se investiguen las opciones de filtro antes que las posibilidades de sanción y que, en la medida de lo posible, las intervenciones de sanción deberían alentar la deliberación adecuada. Sin embargo, la elección racional nos depara una lección más, relativamente evidente. Dado que es probable que existan agentes explícitamente movidos por el interés egoísta en todos los ámbitos de la vida social, es importante que el diseño institucional que se ocupa de cada área contemple sanciones que resulten eficaces como motivaciones para tales personas; sanciones que resulten suficientes para motivar a los canallas en quienes se concentran Mandeville, Hume y sus sucesores.

La idea centrada en el cumplimiento, tal como se presenta en las dos primeras recomendaciones, señala que deberíamos intentar asegurarnos de que el diseño institucional refuerza en los individuos un patrón de conducta que tenga razones independientes y deliberativas para adoptar. Este debe ser nuestro propósito, en lugar de intentar motivarlos a adoptar ese patrón a partir de cero. El principio que agregamos ahora surge del reconocimiento de que es posible que esta idea no se aplique eficazmente a determinadas personas; es posible que no se aplique a los canallas que están más o menos explícita y exclusivamente motivados en su deliberación por el interés egoísta ni a quienes carecen de una inclinación independiente a conducirse de una manera deseable.

No obstante, el principio plantea un problema. Las sanciones que se establecen en apoyo de la deliberación de quienes no son canallas no resultarán adecuadas, generalmente, para controlar a estos últimos. ¿Qué hacer entonces? ¿Cómo podemos asegurarnos de que incluso los canallas sean motivados? Con seguridad, ningún sistema resultará completamente satisfactorio: los canallas nunca estarán completamente contenidos. Sin embargo, para que valga la pena considerar un sistema de sanciones, éste debe imponer ciertas restricciones a quienes se inclinan independientemente al incumplimiento. Debe ser capaz de reducir el perjuicio potencial de la conducta desviada y estar en condiciones de asegurar a quienes cumplen que sus esfuerzos no se vean socavados, explotados o ridiculizados por aquellos que responden a un molde diferente.

El problema es imperioso para quien sigue una estrategia centrada en el cumplimiento. Parecería probable que toda sanción apta para motivar a los canallas menoscabe el tipo de deliberación que mantiene a la mayoría de las personas en la senda correcta. ¿Qué debemos recomendar, entonces? ¿Existe alguna manera de establecer motivadores institucionales que no interfieran con los hábitos de la mayoría?

John Braithwaite ha elaborado un enfoque de la sanción que brinda una respuesta a esta pregunta. La idea es que las sanciones, en especial las penas, pueden diseñarse dentro de una escala progresiva. En el nivel más bajo encontramos las sanciones que se aplican a todos y que alientan idealmente la deliberación. Si las sanciones de ese nivel resultan incapaces de mantener a alguien en la senda correcta –si se descubre que la persona incumple la reglamentación pertinente y demuestra ser, en cierto sentido, un canalla– se recurre a sanciones de un nivel superior, más estricto. El proceso puede continuar en diversos ciclos, avanzando en la jerarquía hacia lo que Braithwaite describe como el «gran garrote» o «el gran revólver» (Ayres y Braithwaite, 1992).

De no contar con esta propuesta de imponer sanciones en una jerarquía progresiva, sería difícil aplicar la estrategia centrada en el cumplimiento que he descrito. El sistema considerado en los dos primeros principios parecería fatalmente vulnerable al perjuicio que pueda causarle un canalla. Resultaría escasamente recomendable para los teóricos de la elección racional por graves que sean los inconvenientes de su alternativa, la estrategia centrada en la desviación. No obstante, contando con la propuesta de establecer una progresión, considero que podemos ser razonablemente optimistas acerca de la estrategia centrada en el cumplimiento. Podemos concentrar nuestra atención en los individuos comunes, como recomienda la estrategia, y tener una conciencia clara de las medidas para tratar con marginales de tendencias canallescas.

Espero que las consideraciones propuestas en esta última sección demuestren por qué la estrategia del diseño institucional que se centra en el cumplimiento debería resultar mucho más atractiva para la teoría de la elección racional que la estrategia más tradicional, centrada en la desviación (es decir, la estrategia de diseñar un mundo concebido para canallas).

La estrategia centrada en el cumplimiento evita limpiamente uno de los problemas de la estrategia centrada en la desviación y, en última instancia, se presenta como la mejor opción. Sin embargo, para concluir existe una última cuestión que me gustaría señalar. He afirmado que la estrategia centrada en la desviación plantea dos problemas. Uno de ellos es que tiende a socavar el cumplimiento espontáneo y el otro –una dificultad más genérica– es que exacerba el problema de controlar a los guardianes, de defenderse de los defensores. La estrategia centrada en el cumplimiento está diseñada explícitamente para tratar el problema del deterioro del cumplimiento. No obstante, ¿cómo funciona con respecto a la otra cuestión?

Por cierto, no exacerba la dificultad tanto como la estrategia centrada en la desviación, ya que no exige un sistema centralizado de sanciones particularmente severas ni necesita otorgar un poder excesivo a un grupo de funcionarios. Por el contrario, en este aspecto, así como respecto al primer problema, la estrategia centrada en el cumplimiento nos predispone hacia maneras de superar la dificultad: nos señala una variedad de medidas a través de las cuales cabría esperar que se mantenga a los guardias en la senda correcta. La estrategia sugiere que deberíamos concentrarnos, en primer lugar, en las medidas de filtro, es decir, en medidas para la selección de individuos y de opciones. También sugiere que pueden existir maneras de sancionar a individuos con autoridad que reforzarían los tipos de deliberación que esperamos de los funcionarios públicos. En especial, que pueden existir maneras de hacerlo que también permitan graduar las sanciones que se aplican a quienes demuestran no ser sensibles a este refuerzo. Sin embargo, debo desistir de explorar esta sugerencia. En este caso, como en otras cuestiones a las que se refiere el presente ensayo, hay terreno para una mayor reflexión e investigación.[13]

Bibliografía

Ayres, Ian and Braithwaite, John. 1992. *Responsive Regulation*. Nueva York, Oxford University Press.

Bardach, Eugene y Kagan, Robert A. 1982. *Going by the Book: The Problem of Regulatory Unreasonableness*. Philadelphia, Temple University Press.

13. Debo agradecer a los participantes en los dos talleres en los que se debatió este material: el primero, en la Universidad ANU (diciembre de 1992); el segundo en Cerisy, Normandía (junio de 1993). Deseo agregar un reconocimiento particular por sus útiles comentarios a John Braithwaite, Geoffrey Brennan, Peter Drahos, John Ferejohn, Bob Goodin, Claus Offe, y una referencia anónima.

Becker, Gary. 1976. *The Economic Approach to Human Behavior*. Chicago, University of Chicago Press.

Braithwaite, John. 1989. *Crime, Shame and Reintegration*. Cambridge, Cambridge University Press.

Braithwaite, John y Pettit, Philip. 1990. *Not Just Deserts: A Republican Theory of Criminal Justice*. Oxford, Oxford University Press.

Braithwaite, John y Makkai, Toni. 1994. «Trust and compliance». *Policing and Society*, 4, pp. 1-12.

Brennan, Geoffrey y Buchanan, James. 1981. «The normative purpose of economic «science»: rediscovery of an eighteenth century method». *International Review of Law and Economics*, 1, pp. 155-166.

Brennan, Geoffrey y Pettit, Philip. 1990. «Unveiling the vote». *British Journal of Political Science*, 20, pp. 311-333.

Brennan, Geoffrey y Pettit, Philip. 1991. «Modeling and motivating academic performance». *Australian Universities' Review*, 34, pp. 4-9.

Brennan, Geoffrey y Pettit, Philip. 1993. «Hands invisible and intangible». *Synthese*, 94, pp. 191-225.

Buchanan, James. 1975. *The Limits of Liberty*. Chicago, University of Chicago Press.

Davidson, Donald. 1984. *Inquiries into Truth and Interpretation*. Oxford, Oxford University Press. [*De la verdad y de la interpretación: Fundamentales contribuciones a la teoría del lenguaje*. Barcelona, Gedisa, 1989.]

Durkheim, Émile. 1961. *Moral Education: A Study in the Theory and Application of the Sociology of Education*, trad. de E. K. Wilson y H. Schnurer. Nueva York, Free Press.

Eells, Ellery. 1982. *Rational Decision and Causality*. Cambridge, Cambridge University Press.

Elster, Jon. 1983. *Sour Grapes*. Cambridge, Cambridge University Press. [*Uvas amargas: sobre la subversión de la racionalidad*. Barcelona, Edicions 62, Península, 1988.]

Elster, Jon (comp.). 1986a. *Rational Choice*. Oxford, Blackwell.

Elster, Jon. 1986b. «The market and the forum: three varieties of political theory», en Jon Elster y A. Hylland (comps.), *Foundations of Social Choice Theory*. Cambridge, Cambridge University Press, pp. 103-128.

Fogel, Robert W. y Stanley L. Engerman. 1974. *Time on the Cross: The Economics of American Negro Slavery*. Boston, Little, Brown. [*Tiempo en la cruz: la economía esclavista en los Estados Unidos*. Madrid, Siglo XXI, 1981.]

Goodin, Robert E. 1992. *Motivating Political Morality*. Oxford, Blackwell.

Hargreaves-Heap, Shaun; Hollis, Martin; Lyons, Bruce; Sugden, Robert y Weale, Albert. 1992. *The Theory of Choice*. Oxford, Blackwell.

Harsanyi, John. 1969. «Rational choice models of behavior versus functionalist and conformist theories». *World Politics*, 22, pp. 513-538.

Heath, Anthony. 1976. *Rational Choice and Social Exchange*. Cambridge, Cambridge University Press.

Hindess, Barry. 1988. *Choice, Rationality and Social Theory*. Londres, Unwin-Hyman.
Holmes, Stephen. 1990. «The secret history of self-interest», en Mansbridge, Jane J. (comp.), *Beyond Self-interest*. Chicago, University of Chicago Press, pp. 267-286.
Hume, David. 1875. «Of the independence of Parliament», en T. H. Green and T. H. Grose (comps.). *Philosophical Works*, Londres, vol. 3.
Lepper, M. R. y Greene, D.. 1978. *The Hidden Costs of Reward*. Hillsdale, N.J., Erlbaum.
Levi, Margaret. 1987. *Of Rules and Revenue*. Berkeley, University of California Press.
Lovejoy, Arthur O. 1961. *Reflections on Human Nature*. Baltimore, Johns Hopkins University Press.
Luce, R. D. y Raiffa, Howard. 1957. *Games and Decisions*. Nueva York, Wiley.
Mandeville, Bernard. 1731. *Free Thoughts on Religion, the Church and National Happiness*. Tercera edición. Londres.
McCubbins, Matthew D. y Schwartz, Thomas. 1984. «Congressional oversight overlooked: police patrols vs. fire alarms». *American Journal of Political Science*, 28, pp. 165-179.
McCullagh, C. Behan. 1991. «How objective interests explain action». *Social Science Information*, 30, p 29-54.
McLean, Iain. 1987. *Public Choice: An Introduction*. Oxford, Blackwell.
Parfit, Derek. 1984. *Reasons and Persons*. Oxford, Oxford University Press.
Peters, Charles y Branch, Taylor. 1972. *Blowing the Whistle: Dissent in the Public Interest*. Nueva York, Praeger.
Pettit, Philip. 1982. «Habermas on truth and justice», en G. H. R. Parkinson, *Marx and Marxisms*. Cambridge, Cambridge University Press, pp. 207-228.
Pettit, Philip. 1985. «The Prisoner's Dilemma and social theory: an overview of some issues». *Politics*, 20, pp. 1-11.
Pettit, Philip. 1992. «Instituting a research ethic: chilling and cautionary tales». Academy of Social Sciences Annual Lecture (1991), University House, Canberra. Reimpreso, con ligeras correcciones, en *Bioethics*, 6 (1992), pp. 89-112 y *Bioethics News*, 11 (4), 1992.
Pettit, Philip. 1993a. *The Common Mind: An Essay on Psychology, Society and Politics*. Nueva York, Oxford University Press.
Pettit, Philip. 1993b. «Normes et choix rationnels«. *Reseaux*, 62, pp. 87-112.
Pettit, Philip. «The virtual reality of *Homo economicus*». *Monist*.
Runciman, W. G. 1972. *Relative Deprivation and Social Justice*. Harmondsworth, Inglaterra, Penguin.
Sen, Amartya. 1970. *Collective Choice and Social Welfare*. Edinburgh, Oliver & Boyd. [*Elección colectiva y bienestar social*. Madrid, Alianza, 1976.]
Sen, Amartya. 1982. *Choice, Welfare and Measurement*. Oxford, Blackwell.
Smith, Adam. 1982. *The Theory of the Moral Sentiments*. D. D. Raphael y A. L. Macfie (comps.). Indianápolis, Ind., Liberty. [*La teoría de los sentimientos morales*. Madrid, Alianza, 1997.]

Taylor, Michael. 1988. «Rationality and collective actino», en Michael Taylor (comp.), *Rationality and Revolution*. Cambridge, Cambridge University Press.
Titmuss, Richard. 1971. *The Gift Relationship*. Londres, Allen & Unwin.
White, Morton. 1987. *Philosophy, the Federalist, and the Constitution*. Nueva York, Oxford University Press.
Williamson, Oliver E. 1983. «Credible commitments: using hostages to support exchange». *American Economic Review*, 71, pp. 519-540.
Wills, Garry (comp.). 1982. *The Federalist Papers*. Nueva York, Bantam Books.
Zimring, Franklin E. y Hawkins, Gordon J. 1973. *Deterrence: The Legal Threat in Crime Control*. Chicago, University of Chicago Press.