



UNIVERSIDAD NACIONAL DE COLOMBIA

# **Castigos y recompensas como mecanismos evolutivos para superar los dilemas sociales**

**GUSTAVO ADOLFO SILVA CARRERO**

**UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS HUMANAS  
DEPARTAMENTO DE FILOSOFÍA  
MAESTRÍA EN FILOSOFÍA**

Bogotá, abril del 2014

# Castigos y recompensas como mecanismos evolutivos para superar los dilemas sociales

TESIS DE GRADO PRESENTADA PARA OPTAR POR EL TÍTULO DE:

*MAGÍSTER EN FILOSOFÍA*

PRESENTADA POR:

GUSTAVO ADOLFO SILVA CARRERO

DIRECTOR: ALEJANDRO ROSAS LÓPEZ. PHD.

Línea de Investigación:

Filosofía de la ciencia, evolución de la cooperación

Grupo de Investigación:

Ética Comportamiento y evolución

(Departamento de Filosofía)

UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS HUMANAS  
DEPARTAMENTO DE FILOSOFÍA  
MAESTRÍA EN FILOSOFÍA

Bogotá, abril del 2014

Dedicado a mi gran amigo Felipe Lanchas,  
quien impulsó con sabiduría  
este escrito inasible y efímero.

# **Castigos y recompensas como mecanismos evolutivos para superar los dilemas sociales**

## **Punishments and rewards as evolutionary mechanisms to overcome social dilemmas**

### Resumen

En este trabajo pretendo solucionar, bajo un marco teórico que asume la racionalidad acotada e imperfecta, el problema paradójico de la evolución de la cooperación, que en últimas es el problema de la explotación de los sistemas de vigilancia que estabilizan la cooperación. En este caso mi propuesta es simple y directa. Postulo que si encontramos un tipo de sanciones y recompensas que entreguen a sus ejecutores beneficios que superen la inversión o el costo de ejecutar dichos incentivos, podemos alcanzar un sistema de cooperación autovigilante y estable. ¿Cómo se logra esto? Concentrándonos en los castigos institucionales que ya Hobbes propone como solución al dilema del contrato social y el seguimiento de leyes en un estado de civilidad. Así pues, afirmo que los incentivos institucionales no deben ser considerados como mecanismos altruistas, ni a nivel biológico, ni a nivel psicológico.

Palabras claves: Filosofía de las ciencias, Cooperación, Castigos, Recompensas, Dilemas sociales, Evolución de la cooperación.

### Abstract

I intend to solve, under a theoretical framework that assumes bounded rationality and imperfect, the paradoxical problem of the evolution of cooperation, which is the problem of exploitation of surveillance systems that stabilize cooperation. My proposal is simple and straightforward. I postulate that if we find a kind of sanctions and rewards, that it delivered his executioners benefits greater than the investment or the cost of implementing such incentives, we can achieve a system of self-monitoring and stable cooperation. How is this accomplished? Focusing on institutional punishments, which Hobbes proposed as a solution to the dilemma of the social contract and monitoring of laws in a state of civility. So I say that the institutional incentives should not be considered altruistic mechanisms or biologically or psychologically.

Keywords: Science philosophy, Cooperation, Punishment, Rewards, Social Dilemmas, Evolution of cooperation.

## TABLA DE CONTENIDO

<b>0. Introducción</b>	<b>6</b>
<b>1. Dilemas sociales: causas y soluciones</b>	<b>12</b>
1.1 <i>Definición y Estructura de los Dilemas Sociales</i>	12
1.2 <i>Factores que Influyen en la Conformación de Situaciones Dilemáticas</i>	17
1.2.1 Principales Factores que Influyen en la Conformación de Situaciones Dilemáticas	20
1.2.2 Limitación en la Información disponible	21
1.2.3 Orientaciones del valor social	23
1.2.4 Grupos grandes	24
1.3 <i>Soluciones a los Dilemas Sociales</i>	24
1.3.1 Soluciones motivacionales	25
1.3.2 Soluciones Estratégicas	28
1.3.3 Soluciones estructurales	28
1.3.4 Soluciones para los Dilemas Sociales	35
<b>2. Cooperación, reciprocidad y castigo. Un enfoque evolutivo</b>	<b>38</b>
2.1 <i>Altruismo parental y selección individual</i>	41
2.2 <i>Altruismo Recíproco</i>	43
2.3 <i>El Dilema del Prisionero Iterado (DIP)</i>	46
2.4 <i>Problemas para la estabilidad de la reciprocidad</i>	51
2.5 <i>Castigo como solución a la inestabilidad de la cooperación por reciprocidad</i>	52
2.6 <i>Dilemas de segundo orden en juegos de bienes públicos y comunales</i>	56
2.7 <i>Selección de grupos como explicación al altruismo</i>	62
2.8 <i>La controvertida Selección de Grupos</i>	64
<b>3. Problemas de coordinación en la teoría política del Estado. Castigos y recompensas</b>	<b>67</b>
3.1 <i>Egoísmo, guerra y racionalidad en Hobbes</i>	68
3.1.1 Inferencia hobbesiana del estado de guerra	69
3.1.2 El dilema del estado de naturaleza	73
3.1.3 Acuerdo para salir del estado de naturaleza	75
3.2 <i>Problemas de coordinación</i>	86
3.3 <i>Sanciones e incentivos. El papel del soberano hobbesiano</i>	90
<b>4. Conclusiones</b>	<b>99</b>
<b>5. Bibliografía</b>	<b>103</b>

## 0. INTRODUCCIÓN

Por décadas los economistas y científicos han estudiado los dilemas sociales, bajo el supuesto de que muchos -por no decir la mayoría- de nuestros problemas sociales tienen que ver con la dificultad de generar comportamientos cooperativos en las situaciones que tales dilemas plantean. En este sentido, gran parte de la investigación desarrollada en teoría de juegos se ha enfocado en encontrar estrategias simples que permitan identificar líneas de acción pertinentes para alcanzar o estabilizar comportamientos cooperativos en interacciones sociales de todo tipo. De la misma forma, el trabajo en psicología social se ha dirigido a desentrañar los mecanismos internos que las personas emplean cuando se enfrentan a situaciones en donde sus intereses particulares están en franca oposición con los intereses del grupo social al que pertenecen.

Una definición más o menos aprobada por la comunidad de científicos sociales afirma que los dilemas sociales son situaciones en las que la racionalidad individual lleva a una irracionalidad colectiva (Kollock, 1998). Así, las decisiones racionales de un individuo en ciertas situaciones pueden ir en contra de los intereses colectivos del grupo social en el que participa (mucho más a menudo de lo que se cree).

El psicólogo Robyn Dawes, en su famoso artículo *Social Dilemmas* de 1980, sostiene que los dilemas sociales se caracterizan por dos simples propiedades, a saber:

- a) Cada individuo recibe un gran pago por una elección socialmente no cooperativa, mientras que recibe uno menor por una socialmente cooperativa.
- b) Pero, si todos los individuos realizan una elección socialmente cooperativa, todos recibirán un mejor pago que si todos realizarán una elección socialmente no cooperativa.

De esto podemos colegir que en las situaciones descritas por los dilemas sociales se presenta una tensión entre los beneficios individuales (racionalidad individual) y los beneficios del grupo (racionalidad colectiva), dado que existen incentivos para que el individuo no actúe como lo requieren los intereses del grupo; provocando, más bien, que la persona se concentre en satisfacer sus intereses particulares e inmediatos (Rosas, 2007, 108).

En general, la solución a los dilemas sociales será, por tanto, la de encontrar dispositivos que aseguren el comportamiento cooperativo en aquellos individuos que interactúan en situaciones

dilemáticas como las ya descritas. En otras palabras, como lo dirían los científicos: es la identificación de ciertos mecanismos que permitan la coordinación de preferencias en situaciones particulares que terminen por beneficiar al grupo en general.

Sin embargo, es generalizada, tanto en la economía, en la teoría de juegos y hasta en la biología evolutiva, la tesis de que la cooperación se constituye en un comportamiento paradójico, pues es irracional, como dirían los teóricos sociales, o nada adaptativa, como dirían los estudiosos de la teoría evolutiva.

Así pues, desde el punto de vista de la teoría económica, los seres humanos nos comportamos como maximizadores egoístas para satisfacer nuestros intereses en situaciones tales como los dilemas sociales. El cálculo racional de las opciones presentadas en este tipo de situaciones nos aconseja no cooperar. De otro lado, vista la cooperación desde la teoría de la evolución encontramos que nuevamente es una anomalía del comportamiento. Esto se debe a que para los evolucionistas el comportamiento más adaptativo en los animales es el egoísta si se tiene en cuenta que el universo biológico es una eterna competencia en donde gana el más fuerte, o el más rápido, o el más feroz, etc. (Darwin, 1859). Así, bajo estos presupuestos, un organismo buscará su propio beneficio a toda costa y aquel que no lo haga condena su linaje a la extinción.

Ahora bien, aunque es cierto que en la biología se admite la existencia de un tipo especial de *altruismo*, que, podría decirse, es el concepto equivalente al de cooperación social en economía, su característica principal es que es un altruismo *sui generis* que en últimas está destinado a beneficiar al organismo que lo ejerce. En otras palabras, el altruismo biológico no es más que un egoísmo bien disimulado.

En este sentido, William Hamilton (1964) sostiene que en ciertos casos, si no en la mayoría, los comportamientos altruistas o cooperativos en los animales pueden ser explicados desde la perspectiva darwiniana de la lógica de la selección individual (la lucha por la existencia). En *The Genetical Evolution of Social Behaviour*, Hamilton propone una teoría que amplía el concepto darwiniano de éxito adaptativo y que incluso mantiene su influencia en teorías posteriores como la del “Altruismo recíproco” de Trivers (1971) o la del “Gen egoísta” de Dawkins (1976). La propuesta de Hamilton (1964) muestra que los individuos pueden mejorar su aptitud (definida por la tasa reproductiva del organismo), ayudando a incrementar la aptitud de parientes cercanos que comparten sus mismos genes. En otras palabras, el altruismo en el plano de la biología es una cooperación con el único fin de sacar ventajas a mediano o largo plazo. Pues, al ayudar a parientes, a expensas de su propio beneficio individual, se asegura la

supervivencia de sus genes, de su familia, de su estirpe.

Con todo, las anomalías de la cooperación identificadas en la teoría evolutiva o la teoría de juegos han provocado que los científicos sostengan que, desde el punto de vista económico, es irracional transferir beneficios a otros en interacciones cooperativas, y maladaptado o “ecológicamente irracional” desde el punto de vista evolutivo (Rosas, 2007).

Empero, y es aquí donde se presenta la paradoja, en nosotros, los humanos (y algunas especies más), podemos identificar sin dificultad situaciones que parecen contradecir las conclusiones antes señaladas. La cooperación se da permanentemente entre personas en interacciones sociales sin que parezcan tener intereses mezquinos. En este sentido, a partir de la década de los ochenta del siglo pasado numerosos experimentos han demostrado que las conclusiones de la teoría de juegos son altamente inexactas, pues al menos el 50% de los jugadores cooperan en dilemas del prisionero de una sola jugada (Rosas, 2007). Asimismo, en juegos de bienes públicos se ha encontrado que la contribución promedio se ubicaba entre el 40% y 60% (Dawes & Thaler, 1988).

Por otro lado, lo que han mostrado los distintos estudios es que los seres humanos nos valemos no solamente de la razón, sino de mecanismos que tienen en cuenta las emociones y que nos involucran con los demás de una manera distinta a lo meramente racional y calculador, haciendo posible conductas cooperativas en ambientes que presentan incentivos para el egoísmo.

Siguiendo a Fehr & Fischbacher (2003) el resultado de estos experimentos nos permite concluir que los seres humanos -a diferencia de lo que propone el supuesto canónico de la economía- de algún modo nos preocupamos por la cooperación y la equidad en nuestras interacciones sociales y que, en algún sentido, somos más “jugadores racionales-emocionales” que meramente racionales.

Algunos teóricos de la evolución han querido resolver esta paradoja de la cooperación acudiendo a un concepto ampliamente controvertido en la misma biología, a saber: la Selección de grupos. Bajo esta propuesta conceptual los científicos intentan mostrar que si se considera a los grupos como organismos individuales en competencia, o unidades de selección, se puede entender mejor por qué al interior de los grupos, manadas, clanes etc., es factible encontrar individuos altruistas que entregan beneficios a los demás, a costa del suyo propio. Esto se explica, porque, como afirman Sober & Wilson (1998), los grupos se fortalecen y son más

competitivos frente a otros grupos cuando en ellos -en su interior-, se presenta una tasa alta de altruistas que “juegan en equipo”. Un grupo con altruistas será más fuerte, más rápido, más feroz, frente a grupos en donde solo existen individuos egoístas que únicamente trabajan por sus propios beneficios y no aportan para ningún objetivo común, como la defensa del grupo. Algunos han alcanzado a sostener que esta teoría de la Selección de grupos tiene su origen en la propia reflexión de Darwin sobre los comportamientos amistosos de individuos en los clanes y grupos primitivos. En su libro *El origen del hombre* de 1871, Darwin postula que estos comportamientos, que aparentemente van en contra de su teoría de la Selección natural, se pueden explicar en tanto que benefician la competencia entre clanes para su supervivencia a una escala mayor.

Mi propuesta en este trabajo es la de encontrar una fórmula menos controvertida que la de la Selección de grupos para explicar la cooperación en los seres humanos. Para eso propondré una solución enmarcada bajo los presupuestos de la teoría de la elección racional, pero teniendo en cuenta una racionalidad influida por las emociones (un tipo de racionalidad acotada) que nos permita coordinar nuestras preferencias para que de esta forma se puedan superar los dilemas sociales. En otras palabras, y desde un punto de vista filosófico, pretendo explicar la posibilidad de la cooperación humana valiéndome de la racionalidad en los individuos sin caer en la paradoja del egoísmo calculador. Así mismo, mi explicación intenta emplear coherentemente los supuestos de la selección individual para mostrar que junto con una racionalidad acotada se puede dar cuenta de la evolución de la cooperación en seres humanos.

Así pues, analizo las posibilidades de la cooperación teniendo en cuenta la influencia de mecanismos tales como las recompensas y los castigos. En este sentido, intentaré explicar cómo los castigos y las recompensas, considerados como mecanismos adaptativos y alimentados por las emociones que ellos producen, transforman la estructura de situaciones dilemáticas como el Dilema del prisionero o la Provisión de bienes públicos, para convertirlas en situaciones en donde la coordinación de preferencias e intereses es francamente posible y, por ende, la conducta cooperativa puede ser estable.

Ahondaré, por lo tanto, en un primer capítulo en la “taxonomía” de los dilemas sociales, identificando su estructura y diversas características para acceder de manera más adecuada a sus posibles soluciones. En este sentido, mostraré que el desarrollo de normas sociales, castigos y recompensas es una de las soluciones más eficientes para buscar la cooperación tanto desde el punto de vista económico como desde el punto de vista evolutivo.

En un segundo capítulo me dedicaré a exponer tres aspectos relacionados con la búsqueda de una explicación evolutiva de la cooperación. a) Reconstruiré la postura, más o menos extendida, de algunos teóricos (Sober & Wilson, 1998; Fehr & Gächter, 2002; Boyd & Richerson, 1992; Frank, 1988) que afirman que estrategias en donde se incluye el castigo ligado al cumplimiento de normas son más eficientes para superar la paradoja de la cooperación. Dicha postura sostiene que el castigo se convierte en un bien público de segundo orden que estabiliza la cooperación en bienes públicos de primer orden. b) Paralelamente, identificaré los problemas y debilidades que esta postura presenta para una interpretación más adecuada de las interacciones sociales que se dan en la realidad. En especial, presentaré la dificultad relacionada con que si los castigos y las recompensas se convierten en bienes públicos, ahora de segundo orden, entonces ellos, consecuentemente, se verán enfrentados a la explotación de los oportunistas; de la misma forma en la que son explotados los bienes públicos de primer orden. c) Por último, expondré los argumentos que apoyan la Selección de grupos, postulada por algunos científicos como la única solución a la anomalía de la cooperación (Sober & Wilson 1998). Esta solución no necesita enfrentarse a las dificultades relacionadas con la explotación de los bienes públicos en cualquiera de sus órdenes, puesto que apela a una fuerza evolutiva a nivel de grupos que presionaría para que se desarrollen comportamientos altruistas dentro de cada grupo o comunidad. Cierro este capítulo elaborando un recuento de posturas a favor y en contra de la Selección de grupos que demuestra con claridad la gran inestabilidad y controversia en la que aún se encuentra sumergida dicha teoría.

En el último y tercer capítulo me concentraré en el análisis de las soluciones propuestas por Jean Hampton (1986) a los problemas que se presentan en la teoría hobbesiana del contrato social. Apoyado en la propuesta de Hampton (1986) y que se construye a partir de la teoría de juegos, quiero proponer una salida a los problemas ya expuestos en el segundo capítulo de este trabajo, que tienen que ver con que un bien público de segundo orden está expuesto también a la explotación indebida.

En este sentido, quiero mostrar que si se tienen en cuenta un tipo de recompensas y castigos muy poco abordados en la literatura especializada, a saber: los castigos y las recompensas institucionales, existen buenas razones para considerar que este tipo de mecanismos pueden evitar caer en una explotación como bien público de segundo orden y eliminar así, la posibilidad de que los castigos y las recompensas se conviertan también en un dilema social. Además, este tipo de castigos y recompensas tiene el poder de modificar la estructura de los dilemas, fomentando la coordinación de preferencias de individuos para la cooperación. Tanto la

revisión del papel de los castigos y las recompensas en los dilemas, como la coordinación de las preferencias para el fomento de la cooperación se mantienen bajo los presupuestos de la teoría de la elección racional, pero que a su vez presenta un tipo de racionalidad acotada por la fuerte influencia de las emociones producidas por los mecanismos evolutivos del castigo y la recompensa.

Finalmente, en las conclusiones, elaboraré una recapitulación muy general de lo expuesto para dar más claridad a mi propuesta, a saber: que la evolución de la cooperación humana puede ser explicada apelando a los recursos de la racionalidad, sin caer en la paradoja del egoísmo universal de los seres racionales. De tal forma que se evitará acudir a explicaciones altamente controvertidas como las de la Selección de grupos, en donde aún no hay definición sobre los presupuestos y condiciones epistemológicas y ontológicas de dicha teoría (Nesse, 1994; Andrade & Fajardo, 2008).

# 1. DILEMAS SOCIALES: CAUSAS Y SOLUCIONES

La especie humana, desde muy temprano en su evolución, se ha caracterizado por asumir de forma asociativa y mancomunada los retos que la naturaleza pone en su camino. Sin embargo, también desde muy temprano, los seres humanos han experimentado una gran tensión entre su instinto competitivo y altamente individualista y dicha forma gregaria de alcanzar objetivos mayores en comunidad.

Esta lucha interna entre el ser egoísta que se dirige a alcanzar el éxito personal y aquel individuo que sacrifica, en alguna medida, ese éxito para aportar en una empresa común, se ve reflejada significativamente en aquellas situaciones dilemáticas de la vida diaria en donde la racionalidad individual obstaculiza el bienestar de una comunidad. Estos casos, nada excepcionales en la realidad, son denominadas *dilemas sociales* por los investigadores. Como ya se sugirió, los dilemas sociales son situaciones en las que la racionalidad individual lleva a una irracionalidad colectiva (Kollock, 1998, 183). En otras palabras, las decisiones racionales de un individuo en ciertas situaciones pueden ir en contra de los intereses colectivos del grupo social en el que participa; por tanto, esta racionalidad individual se encarna en una irracionalidad colectiva.

## ***1.1 Definición y Estructura de los Dilemas Sociales***

Los estudios sobre este dominio de las ciencias sociales han determinado que los dilemas sociales se caracterizan por una especial estructura de pagos, pues son situaciones en las que un individuo recibe por una elección socialmente no cooperativa o egoísta un buen pago. Pero, si todos los individuos realizan una elección socialmente cooperativa, todos recibirán un mejor pago que si todos realizarán una elección socialmente no cooperativa (Dawes, 1980, 170).

Con todo, las situaciones dilemáticas de este tipo son producidas, comúnmente, por factores externos al individuo que influyen en su conducta, incentivándolo para que tome elecciones y decisiones individuales en desmedro de las elecciones y decisiones colectivas que serán siempre benéficas si todos o una buena parte de los individuos las asumen, en lugar de que todos desarrollen actitudes egoístas. En este sentido, los dilemas sociales son escenarios en los que a las personas se les presenta un claro y nada ambiguo incentivo para llevar a cabo una elección egoísta que, cuando se realiza por todos los miembros de un grupo o comunidad, conduce a resultados poco deseados y pobres con respecto a aquellos resultados que hubieran podido ser

alcanzados si todos los individuos actuaran de forma distinta, es decir racionales colectivamente (Dawes, Messick, 2000, 112).

Las situaciones que caen bajo la etiqueta de dilemas sociales son, entre otras: el calentamiento climático del planeta, la deforestación, la depredación silvestre, la pesca indiscriminada, la explosión demográfica, la evasión de impuestos, el abstencionismo en las votaciones, etc.

Por décadas la investigación en torno a los dilemas sociales ha utilizado prolíficamente la teoría de juegos, un dominio de las matemáticas esbozado a finales de los años cuarenta por el matemático húngaro-estadounidense John von Neumann. La economía, la psicología social, las ciencias políticas y hasta la biología se han servido de las herramientas forjadas en la teoría de juegos para analizar y ensanchar nuestro conocimiento de los dilemas sociales.

Desde este punto de vista, los dilemas sociales son juegos caracterizados por presentar uno o más equilibrios de Nash<sup>1</sup> que no son eficientes en el sentido del Optimo de Pareto<sup>2</sup>. En otras palabras, los dilemas sociales se caracterizan porque el Equilibrio de Nash que define racionalmente la mejor forma de actuar y, por tanto, determina la estrategia dominante<sup>3</sup> del juego, no coincide con la situación Pareto Eficiente del juego. Es decir, no coincide con la mejor situación -posible en el juego- en donde no hay otras opciones que permitan incrementar los beneficios de uno o más participantes sin que se vean reducidos los beneficios de los demás. Así pues, como sostiene Dawes (1980), un dilema social es provocado porque todos los participantes en una situación presentan una estrategia dominante que genera un resultado Pareto deficiente (Dawes, 1980, 179).

Así pues, desde la teoría de juegos se ha intentado, con relativo éxito, describir las situaciones dilemáticas de la vida diaria a través de múltiples modelos que pueden dividirse principalmente en dos categorías, a saber: i) aquellos en donde se tiene en cuenta la interacción entre dos personas y, ii) aquellos en los que se atiende a la interacción de tres o más personas, denominados juegos de N-personas o multi-agente.

---

<sup>1</sup> Se denomina Equilibrio de Nash a una combinación de estrategias (una por cada jugador) tal que ningún jugador aumente sus ganancias por cambiar su estrategia unilateralmente. En otras palabras, un equilibrio de Nash se da cuando la elección estratégica de cada jugador es la respuesta óptima a la elección estratégica de los demás jugadores.

<sup>2</sup> Una situación es *Óptima de Pareto* o *Pareto Eficiente* si no existe otra opción en la que se pueda mejorar el pago o los beneficios de una persona sin disminuir el pago o los beneficios de las demás personas que intervienen en la interacción. En este mismo sentido, una situación *Pareto Deficiente*, es aquella en la cual existe al menos una forma distinta de mejorar el pago de uno o más participantes sin reducir el beneficio de los demás.

<sup>3</sup> Una estrategia dominante puede definirse como aquella que estipula un curso de acción para un individuo que resulta ser la mejor opción (dar el mejor resultado) independientemente de las decisiones de los demás.

Los juegos más estudiados de la primera categoría son el Dilema del prisionero (DP) -iterado o no-, El Dictador, el Juego de la gallina, la Batalla de los sexos y el Juego de Confianza (*Assurance Game*). Los primeros dos han sido clasificados dentro de una categoría general denominada: juegos de cooperación que refleja los problemas patentes en la cooperación de la racionalidad individual y las preferencias egoístas de los individuos. Los tres últimos, son juegos que expresan problemas de coordinación de preferencias, teniendo en cuenta los presupuestos de la teoría de la elección racional. En este caso, las preferencias pueden definirse dependiendo de las preferencias de los otros. Desarrollaré una caracterización más detallada de los juegos de cooperación y los problemas de coordinación de preferencias en la Sección 3.2. Ahora, para respetar el orden de la argumentación, quiero concentrarme en la explicación del DP; básico para comprender las dinámicas representadas por la teoría de juegos.

El DP fue desarrollado en 1950 por los matemáticos Merrill Flood y Melvin Dresher. Este es un juego de suma no-cero, es decir un juego en el que existe la posibilidad de que todos los participantes ganen algo o, en sentido estricto, de que ningún participante pierda, de tal suerte que la colaboración entre participantes se puede dar. En los juegos de suma cero, no puede presentarse ningún tipo de colaboración, dado que necesariamente la ganancia de un jugador es la pérdida del otro.

La versión original del DP tiene como característica principal el hecho de que la interacción entre dos participantes se da una única vez y, además, la recompensa por desertar (D) mientras el otro coopera (C) es más alta que la de cooperar sin importar qué haga el otro participante. La estructura de preferencias que este dilema presenta es la siguiente:  $DC > CC > DD > CD$ . Bajo el marco de la teoría de juegos, el DP presenta una estrategia dominante que es la de siempre desertar o no cooperar, dado que, sin importar la elección del otro jugador, existen más probabilidades de ganar algo o no perder si no se coopera. Por otro lado, se corren más riesgos si se coopera, pues, asumiendo la racionalidad de la contraparte, será muy probable que el individuo deserte y, por lo tanto, explote la conducta cooperativa.

Tabla 1.1 Matriz de pagos del DP.

		Jugador A	
		C	D
Jugador B	C	(2)	(5)
	D	(0)	(1)

En la tabla 1.1 podemos ver con mayor detalle las opciones que a los jugadores se les presenta en un DP. Por ejemplo, si el jugador A (columnas) elige no cooperar o desertar (D), entonces el jugador B (filas) tendrá dos opciones, cooperar o no, pero la que mayor pago le reporta es la de no cooperar (D)=1, así que preferirá desertar en esta situación. Ahora bien, si suponemos que el jugador A elige cooperar (C), nuevamente el jugador B tiene las mismas dos opciones y, nuevamente, no cooperar (D)=5 es la elección que mayor recompensa le reporta en esta situación particular. Así pues, se ve cómo la estrategia dominante en el juego será siempre desertar, porque sin importar la elección del contrincante la mejor opción es no cooperar. De lo que se sigue que la situación en donde ambos jugadores no cooperan es un Equilibrio de Nash, pues su estrategia dominante los llevará a dicho estado de cosas. Ahora bien, en aquellos casos en donde ambos jugadores desertan no es el escenario Pareto Eficiente del juego, puesto que existe un resultado en donde pueden mejorarse los beneficios de ambos, a saber, la de que juntos cooperen. Esta última es el resultado Pareto Eficiente que evidentemente no coincide con el Equilibrio de Nash y, por tanto, esto provoca que dicho equilibrio sea deficiente.

Para el caso de los dilemas de N-personas, Kollock (1998) realiza, a su vez, una doble clasificación, a saber:

- a) Limitación social (*Social Fence*): en donde al individuo se le presenta un inmediato y particular costo, pero que representa beneficio para todos los integrantes de una comunidad. En este caso el individuo tiene el incentivo de evitar el costo inmediato, pero

si todos hicieran caso a este incentivo y evitaran el costo, la ganancia para cada uno sería la peor posible. En esta clase de juegos se pueden enmarcar los denominados como *Provisión de Bienes Públicos*.

- b) Trampa social (*Social Trap*)<sup>4</sup>: aquí el individuo se enfrenta a una situación que le reporta beneficios inmediatos, pero que a su vez representa un costo que ha de repartirse entre todos los integrantes de una comunidad o grupo social. En este caso, si todos los individuos sucumben a la tentación del beneficio inmediato, el resultado será un desastre colectivo. Bajo este tipo de juegos se hallan los relacionados con la *Tragedia de los Comunes*.

Parece ser que estos ejemplos de dilemas de N-personas son más adecuados para reflejar las situaciones a las que se enfrentan los individuos en la realidad. Así pues, la provisión de bienes públicos se caracteriza por la imposibilidad de excluir (carácter de *no-exclusión*) a los participantes del usufructo del bien y, además, presenta un carácter de *no-rivalidad*, que se expresa en el hecho de que aunque una persona utiliza el bien público, esto no implica que se disminuya su disponibilidad para otras personas (Kollock, 1998, 189). Como ejemplo de lo anterior puede considerarse la construcción de una alameda en la localidad. Aunque existan vecinos que no hayan colaborado en su construcción, ya sea aportando recursos o fuerza de trabajo, no se los puede excluir de su disfrute y, además, el hecho de que estas personas utilicen la alameda no disminuye su disponibilidad para que la disfruten los otros vecinos.

Los beneficios reportados a los individuos en el dilema de bienes públicos deben verse como una relación entre el nivel de recursos contribuidos y la producción del bien público (Kollock 1998, 189). Así pues, esta relación se comporta de acuerdo a una *Función de producción* que según Kollock (1998) puede ser de cuatro tipos:

1. *Función de producción desacelerada*: en donde la contribución inicial tiene un gran efecto sobre la provisión del bien público, pero contribuciones adicionales encuentran un retorno cada vez menor de los beneficios.
2. *Función de producción lineal*: cada unidad de recursos produce una unidad de retorno o beneficio.

---

<sup>4</sup> El concepto de *Trampa Social* fue propuesto por primera vez en 1973 por el psicólogo J. Platt en su artículo *Social Traps*.

3. *Función de producción acelerada*: al inicio la contribución representa poco beneficio, pero en la medida en que se incrementa la contribución se incrementa exponencialmente el retorno.
4. *Función de producción discontinua*: en este caso el retorno de los beneficios está condicionado por ciertos umbrales o puntos de provisión.

En el contexto de los dilemas comunales o *tragedia de los comunes* la situación es un poco diferente, puesto que aunque comparte con los dilemas de bienes públicos el carácter de *no-exclusión*, no pasa lo mismo con el carácter de *no-rivalidad*. En este caso el uso de un bien por parte de una persona afecta la disponibilidad del bien para los demás. La pesca indiscriminada es un buen ejemplo de lo anterior.

Además, en los dilemas comunales no se presenta ningún tipo de función de producción, por el contrario lo que aquí se debe tener en cuenta es una relación entre el nivel de explotación o uso del bien y la tasa de reabastecimiento o recuperación de dicho bien (Kollock, 1998, 191).

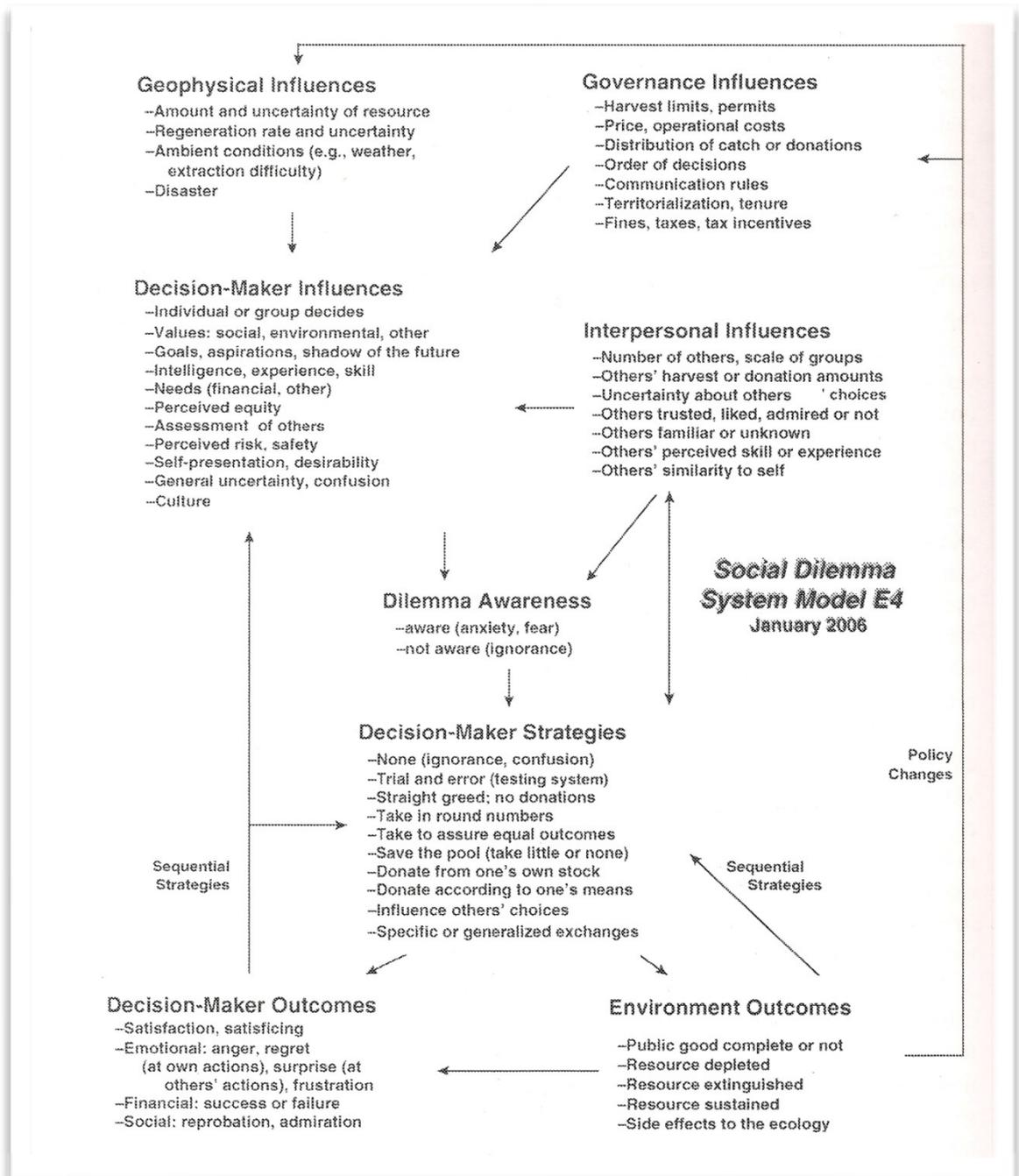
## ***1.2 Factores que Influyen en la Conformación de Situaciones Dilemáticas***

La estrategia dominante que desemboca en un equilibrio deficiente se genera, en las situaciones de la vida real, gracias a la influencia de múltiples factores. Muchas de las investigaciones sobre dilemas sociales se han concentrado en unos pocos factores que pueden considerarse como los más claves para la explicación de estas situaciones dilemáticas. Sin embargo, el psicólogo Robert Gifford de la Universidad de Victoria, Canadá, ha intentado elaborar un modelo que contempla una gran variedad de factores. Su modelo relaciona ocho diferentes categorías que contienen, a su vez, diversas fuentes de influencia que provocan equilibrios deficientes en interacciones sociales del tipo de juegos de N-personas (Gifford, 2008, 267).

En la Figura 1.1 se muestra el modelo propuesto por Gifford, extractado de su artículo *Toward a comprehensive model of social dilemmas* de 2008. Dicho modelo alcanza una complejidad no tenida en cuenta por las investigaciones previas y, sobre todo, expresa muy bien el entramado de dependencias e influencias entre uno y otro factor para provocar una situación dilemática. Con todo, considero que algunos de los factores señalados por Gifford pueden convertirse en triviales, si se quiere implementar a partir de ellos estrategias de solución para los dilemas sociales. Particularmente la primera categoría enunciada por Gifford, *Influencias Geofísicas* (dificultad, entre otras, de la explotación y cuantificación de un recurso por influencias

ambientales), no nos da muchas herramientas para construir o proponer soluciones a situaciones dilemáticas.

Figura 1.1 Categorías de los factores que influyen en la generación de dilemas sociales. Modelo de Gifford.



Para Gifford el valor del modelo radica en que debe postular relaciones claves entre las influencias identificadas y, también, debe ayudar a presentar un sistema complejo de modo comprensible. En este sentido, el modelo de Gifford representa una complejidad, no solamente en cuanto a la agrupación de factores que pueden generar dilemas sociales, sino que además representa, de buena manera, la compleja red de relaciones entre cada una de estas categorías.

Es evidente que los factores que intervienen en la generación de equilibrios deficientes y, por ende, en la aparición de dilemas sociales se pueden relacionar entre ellos de diversas formas. Así que el intento de una clasificación, por más general o detallada que sea, tiene la dificultad de representar factores dinámicos que bien pueden intervenir en diversas categorías de un mismo modelo. Esto se debe a la compleja trama de relaciones que se teje entre ellos y que hace que un dilema social no se produzca por la influencia de un único y simple factor.

Gifford asegura que su modelo puede subsumirse a cuatro grandes agrupaciones de factores que intervienen en un dilema social, a saber: i) aquellos asociados con los recursos mismos (su ausencia, su tasa de regeneración, etc.); ii) aquellos asociados con los propios individuos (sus valores o experiencias, etc.); iii) los relacionados con la interacción entre los individuos (confianza, comunicación, etc.) y iv) los que tienen que ver con la estructura de los dilemas (las reglas que gobiernan la distribución de pagos, etc.). Bajo estos cuatro grandes grupos Gifford ha desarrollado un modelo más detallado que tiene en cuenta alrededor de treinta factores específicos que pueden intervenir en la generación de dilemas sociales (Figura 1.1).

En lo que sigue, adoptando la tendencia de la mayoría de investigaciones (Kollock, 1998; Dawes, 1980; Orbell et al, 1990), me concentraré en el análisis de algunas causas o influencias que considero claves para la conformación de dilemas sociales y necesarias para tener en cuenta al intentar proponer soluciones. En este sentido, intentaré realizar una clasificación más sencilla de los factores implicados en la conformación de situaciones dilemáticas, que tome en cuenta las causas determinantes y esenciales que deben intervenir para llevar a cabo propuestas exitosas de solución. Debo aclarar, por tanto, que mi propuesta de clasificación no pretende reñir con la de Gifford, más completa y compleja, sino que solamente pretende concentrarse en aquellos factores básicos que pueden provocar situaciones dilemáticas y que mediante estrategias dirigidas a ellos puede alcanzarse, con cierto grado de éxito, una superación del dilema social. Mi clasificación, tanto de los factores que intervienen en los dilemas sociales como de las soluciones a dichas situaciones, aunque sencilla, como se verá, expresa de una u otra forma la complejidad de las relaciones entre cada uno de los factores, puesto que en ocasiones su influencia mutua puede hacer que su ubicación en las categorías de clasificación no sea clara y definitiva.

### **1.2.1 Principales Factores que Influyen en la Conformación de Situaciones Dilemáticas**

En el siguiente esquema presento la clasificación de los factores o causas que intervienen en la generación de dilemas sociales, y que considero son los más relevantes para la búsqueda de soluciones a dichas situaciones dilemáticas.

#### **Limitación en la Información disponible**

- Incertidumbre frente a las posibilidades de elección.
- Limitación en la comprensión de las utilidades de una situación específica.
- Incertidumbre sobre interacciones previas de los demás.
- Desconfianza.
- Imposibilidad de generar afecto o empatía.
- Miedo de actuar como un tonto (*sucker*).
- Poca o nula comunicación.
  - o Dificultad en la realización de acuerdos explícitos.
  - o Carencia de discusión entre participantes sobre la estructura del problema y sus posibles salidas y consecuencias.
  - o Dificultad de la aparición de identidad de grupo.

#### **Orientaciones del valor social**

- Aversión a la pérdida.
- Explotación del carácter competitivo de las personas.
- Explotación del carácter individualista de las personas.

## Grupos grandes

- Anonimato.
  - o *Free rider*.
- Baja percepción de eficacia.
- Limitación en la repetición de las interacciones.
- Dificultad en la aparición de la reciprocidad

A continuación revisaremos con más detalle cada factor y su relación con los demás.

### 1.2.2 Limitación en la Información disponible

La limitación en la información disponible provoca una gran incertidumbre frente a las posibilidades de elección en una situación dada, de tal manera que los individuos preferirán siempre actuar guiados por sus intereses particulares, en lugar de arriesgarse a colaborar con los demás jugadores. Sin información o con muy poca las personas pueden estar inhabilitadas para comprender a cabalidad las utilidades de una situación específica o la relación de dichas utilidades con las de los demás (Dawes, 1980, 190), provocando de esta forma que los individuos adopten un comportamiento eminentemente egoísta para asegurar sus propios beneficios.

De otro lado, la incertidumbre puede tener un efecto devastador en la *confianza* de los individuos, no solamente porque no pueden determinar con exactitud la situación en la que se encuentran y, por lo tanto, no saben con exactitud cómo proceder, sino que también, al no poder acceder a la historia de los otros jugadores, al no tener información sobre interacciones previas de los demás, se genera una incertidumbre sobre el carácter del otro como posible cooperadores o posible egoísta (Rosas, 2007, 109). Esto, en otras palabras, tiene que ver con que la incertidumbre hace que los individuos se comporten guiados por el deseo de evitar ser el tonto (*sucker*) en una situación difusa (Dawes, 1980, 187). Pero, además esta limitación en la información imposibilita la generación de cualquier tipo de *afecto*, *empatía* o *confianza* hacia la contraparte que pueda fomentar un comportamiento cooperativo (Dawes 1980, 177).

Algunos estudios han demostrado que la cooperación es eminentemente condicional y esto depende, en buena medida, de la información que puedan tener las personas sobre la

interacción (Bicchieri, 2006), puesto que las personas cooperan siempre y cuando dos tipos de expectativas se vean satisfechas, a saber: las expectativas empíricas, relacionadas con saber o tener la percepción de que existe un suficiente número de personas que siguen una norma social y cooperan en una situación específica y, las expectativas normativas, que tiene que ver con saber o tener la percepción de que hay personas que esperan que uno mismo siga dicha norma en una situación específica (Bicchieri, 2006).

De otro lado, múltiples estudios han demostrado que debido a la poca o nula existencia de *comunicación* en una interacción social, los individuos están más inclinados a comportarse de forma racionalmente egoísta (Rapaport et al, 1962, Bixenstein et al, 1966, Jerdee & Rosen, 1974, Brechner, 1977, Dawes et al, 1977, Edney & Harper, 1978, Liebrand, 1984; Orbell et al, 1990). Esto puede deberse, en primera instancia, a la presencia de una *limitación en la información*, como ya se mencionó. Es decir, que la falta de comunicación no permite que los individuos adquieran información directa de sus contrapartes, provocando que los jugadores consideren a los otros participantes como personas que, en principio, guían su comportamiento por intereses egoístas. Al no presentarse comunicación, las personas no pueden obtener información previa de las intenciones y actitudes de los otros, generándose de esta forma una gran *incertidumbre* y *desconfianza* en torno a las decisiones de los demás, imposibilitando también la realización de acuerdos explícitos para la mutua colaboración.

En segunda instancia, la falta de comunicación sitúa a los individuos en un escenario dilemático, pues sin comunicación las personas pueden no percatarse de la situación en la que se encuentran. En otras palabras, sin comunicación se omite la discusión entre participantes sobre la estructura del problema y sus posibles salidas y consecuencias. Los individuos no saben si pueden conformar grupos para cooperar o si se encuentran en una situación en donde la “ley de la selva” puede ser más efectiva, porque no saben con certeza en qué situación están, no saben qué piensan los otros de la situación en la que se encuentran y qué procedimientos emplearán para salir bien librados.

Por último, la falta de comunicación puede dificultar la adquisición o el refuerzo de identidad de grupo (lo explico con más detalle en la sección 1.3.1.3), mecanismo importante para la generación de cooperación en dilemas sociales.

### 1.2.3 Orientaciones del valor social

Este factor se relaciona con las diferencias en las preferencias de distribución y valoración que asignamos en los resultados de nuestras interacciones frente a los resultados de las interacciones de otros (Kuhlman & Marshello, 1975; McClintock & Liebrand, 1988; Dawes & Messick, 2000; Eek & Gärling, 2008). En otras palabras, los seres humanos asumimos valoraciones distintas de los mismos resultados dependiendo de la forma como se obtienen o de quien los obtenga. Bajo este marco se encuentra el famoso concepto de aversión a la pérdida, en cuanto a que experimentamos una mayor sensibilidad a sufrir costos que a gozar de beneficios (Kahneman & Tversky, 2000).

Los individuos presentan diferencias en su disposición para cooperar o desertar en interacciones sociales, debido a la diversidad de valores que cada uno asigna a las acciones, resultados y cosas del mundo (Biel, 2008, 2). Dependiendo de las diversas orientaciones del valor social y de las consecuentes decisiones realizadas por los individuos, podemos identificar su actitud en alguna de las cuatro clases de jugadores propuestas por Kollock (1998), a saber:

1. *Cooperador*: es aquel que dirige su comportamiento para maximizar los resultados de todas las personas (incluido él) involucradas en una empresa mancomunada. Generalmente los cooperadores son condicionales (Bicchier, 20), pues aunque su acción está guiada por la maximización de resultados en empresas comunes, dicho comportamiento no se da sí no existe alguna certidumbre de que los demás actuaran igual o le retribuirán en subsiguientes interacciones.
2. *Competidor*: son aquellos individuos que desean maximizar la diferencia relativa entre los resultados propios y los resultados de su contraparte. Evidentemente, siempre buscando aventajar al contrario en esta diferencia. En esta clase pueden ubicarse los individuos codiciosos que pretenden conseguir los mejores resultados a costa de los resultados del otros.
3. *Altruista*: el menos probable. Se refiere al individuo que dirige su comportamiento con el fin de maximizar el resultado de la contraparte, incluso en detrimento de su propio resultado.

4. *Individualista*: son los individuos que se comportan de tal manera que maximizan su propio resultado sin importar el resultado de los demás. En esta última clase de jugadores pueden ubicarse los denominados *Free riders* que son personas oportunistas que reciben beneficios explotando la cooperación de los demás en una empresa común.

#### **1.2.4 Grupos grandes**

El tamaño de una comunidad o grupo influye determinadamente en la generación de situaciones dilemáticas (Rapaport et al, 1962; Bixenstein et al, 1966; Marwell & Schimidt, 1972; Hambuerger et al, 1975; Bonacich et al, 1976; Fox & Guyer, 1977; Komorita & Lapworth, 1982; Dawes 1980; Kollock, 1998), ya que grupos grandes permiten con mayor facilidad a los individuos desertar o tener actitudes egoístas sin ser detectados o reconocidos por los demás miembros de la comunidad. Una actitud de *Free rider* puede ser incentivada por el carácter de anonimato que tiene las personas en un grupo extenso.

La percepción de eficacia de las acciones de los individuos se puede ver afectada cuando los grupos o comunidades son grandes, dado que las personas en estas situaciones pueden considerar que sus aportes no serán críticos o cruciales para obtener el bien deseado por la comunidad (no afectarán los resultados) (Kollock, 1998, 200), de tal manera que se abstendrán de colaborar y en ocasiones preferirán guiar su conducta por intereses particulares.

Por otro lado, en grupos grandes la posibilidad de la reciprocidad se desdibuja, dado que la extensión del grupo, el gran número de integrantes, hace poco factible una repetición de interacciones entre los mismos individuos (Axelrod, 1984). En este sentido, gracias a la tendencia de hallar interacciones de una sola jugada o, por lo menos, limitadas en su repetición, las personas encuentran una fuerte inclinación a guiarse por los intereses racionalmente egoístas y “sacar lo suyo” dadas las pobres posibilidades para la reciprocidad.

### ***1.3 Soluciones a los Dilemas Sociales***

Hemos visto los factores más importantes que pueden generar equilibrios deficientes y, por lo tanto, situaciones dilemáticas en la vida diaria. Ahora bien, dichos factores pueden ser inhibidos, disminuidos o transformados para lograr, con relativo éxito, una coordinación de las preferencias de los individuos en situaciones particulares que terminen por beneficiar al grupo en general. En otras palabras, al buscar la estabilidad del equilibrio de las preferencias en situaciones dilemáticas -es decir, al intentar transformar la estructura del juego, pues pasa de

ser un juego con una estrategia dominante a uno de coordinación- estas situaciones se convierten en escenarios de cooperación en donde la racionalidad colectiva guía los comportamientos individuales de los participantes.

Kollock (1998) ha propuesto una buena clasificación de las soluciones a los dilemas sociales, dividiéndolas en tres amplias categorías. La primera, denominada: *Soluciones motivacionales*, se concentra en las soluciones que pueden inhibir los factores de orientaciones diversas del valor social y provocar *confianza, conciencia, moralidad e identidad de grupo*, al valerse de la comunicación. La segunda categoría propuesta se enfoca en las *Soluciones estratégicas* que se concentran en cómo los agentes eminentemente egoístas pueden conformar resultados favorables a partir del comportamiento de otros. Al decir de Kollock este tipo de soluciones no modifica la estructura del juego y principalmente se ven representadas por la iteración en DP<sup>5</sup>. En esta categoría se pretende fortalecer la *reciprocidad* –o como lo diría Dawes (1980) el *Altruismo táctico*-, la *elección discriminada de contrapartes* y la conformación de *compromisos* a través de la comunicación. Por último, Kollock propone la categoría: *Soluciones estructurales*, en donde se encuentran expuestas las estrategias que pretenden modificar la estructura de los juegos dilemáticos, de tal forma que, al convertirse en otro tipo de juegos, dejen de ser situaciones de dilemas sociales. Bajo esta perspectiva, Kollock tiene en cuenta aspectos tales como la *eliminación del anonimato*, la modificación en la *estructura de pagos*, el incremento en la *percepción de eficacia*, el *tamaño del grupo*, la *limitación de las elecciones de los individuos*, las *sanciones y recompensas*, etc.

Aunque considero que la clasificación de soluciones de Kollock es la más incluyente de todas las expuestas en la literatura, insertaré, en donde corresponda, algunas estrategias no tenidas en cuenta por el sociólogo. Abordemos, pues, con mayor detalle estas categorías.

### **1.3.1 Soluciones motivacionales**

#### 1.3.1.1 Modificación en las orientaciones de valor social

La motivación de los participantes en una situación dilemática puede fácilmente transformar la estructura de dicha situación. Sin embargo, como vimos, los individuos asignan valores diversos a las situaciones, acciones y cosas del mundo (Biel, 2008, 2), generando de esta manera una gran diversidad de preferencias alrededor de un resultado. El problema a resolver es ¿cómo hacer

---

<sup>5</sup> Más adelante discutiré la inclusión de la iteración en este grupo de soluciones.

para que las preferencias de los individuos competidores o individualistas se transformen, de tal suerte que los convierta en personas cooperadoras o altruistas?

La modificación de las orientaciones del valor social solo se puede llevar a cabo a través de la educación del individuo. En este caso, Kollock sostiene que la solución a este factor problemático debe generarse desde la escuela. Los niños deben aprender a cooperar y trabajar en grupo antes de competir y de generar actitudes individualistas. Se debe, por tanto, investigar más en torno a la formación de los valores sociales para intentar estrategias de solución a este factor en situaciones dilemáticas (Kollock 1998, 193)

#### 1.3.1.2 Fomento en la comunicación

Como lo expuse previamente, la carencia o limitación de la comunicación es uno de los factores más importantes en la conformación de dilemas sociales. Además, bajo este factor pueden encontrarse otros que influyen determinadamente en la cooperación, tales como, la confianza o los compromisos.

Siguiendo a Messick & Brewer (1983), puede considerarse que la comunicación fomenta la cooperación al *incrementar el conocimiento de los otros* y de sus posibles actitudes y decisiones, facilitando de esta forma la toma de decisiones propia desde un punto de vista colectivo. De otra parte, gracias a la comunicación los individuos pueden alcanzar *acuerdos* o *compromisos* que garanticen la colaboración en empresas comunales y la sanción de aquellas actitudes no cooperativas. La *confianza* intersubjetiva se ve ampliamente reforzada por la generación y el incremento de la comunicación, pues los participantes pueden acceder a las historias de vida de sus contrapartes, identificar gustos, objetivos y actitudes compartidas, etc. Asimismo, la comunicación abre las posibilidades para que se den *situaciones morales*, en donde los participantes de una interacción pueden acordar guiar su comportamiento apelando a lo que es “correcto” o “apropiado”. Mediante diversos estudios (Orbell et al, 1990; Dawes et al, 1977) se ha demostrado que la apelación a principios morales puede generar comportamientos cooperativos. De la misma forma, la comunicación puede crear y reforzar *identidad de grupo* (Dawes, 1991). Es sabido que la identidad de grupo incrementa la cooperación en individuos, pues hace que estos dejen de actuar mediante intereses racionalmente egoístas y se comprometan profundamente con el logro de los objetivos comunes (Brewer & Kramer, 1986; Kollock, 1998; Messick, 1999; Dawes & Messick, 2000).

Un par de aspectos no tenidos en cuenta por Kollock tienen que ver, primero: con que la comunicación genera en el individuo la *humanización* de su contrincante. La comunicación le pone rostro al adversario, pero además, sentimientos con los que pueden identificarse los jugadores (Dawes 1980, 186). El segundo aspecto se relaciona con la posibilidad de discutir sobre la situación en la que se encuentran los individuos. Esta reflexión conjunta sobre el problema puede permitirles aclarar su situación actual y sus posibilidades hacia la búsqueda de soluciones de grupo (Dawes 1980, 186).

#### 1.3.1.3 Identidad de Grupo

Aunque la comunicación puede crear y fomentar la identidad de grupo, existe la posibilidad de que esta identidad se presente en ausencia de comunicación; los individuos pueden disponer de alguna información indirecta de que integran un grupo con otras personas. Empero, aun en ausencia de comunicación la identidad de grupo influye profundamente en la generación de comportamientos cooperativos. En este sentido Dawes & Messick (2000), sostiene que cambios relativamente menores en el ambiente, como la identificación de un individuo como miembro de un grupo, pueden provocar cambios mayores en la toma de decisiones de dicho individuo, inclinándolo su comportamiento hacia la colaboración al interior de su grupo. Estos autores han demostrado que es relativamente fácil conformar un grupo y, así mismo, relativamente fácil adquirir identidad de grupo. Según Dawes & Messick (2000), la gente es capaz de conformar grupos a partir de preferencias estéticas o aun a partir de una percepción arbitraria. Estos autores sostienen que los individuos son más proclives a cooperar con personas que integran su propio grupo que con personas que integran otros grupos (véase también la teoría de conjuntos evolutiva propuesta por Nowak & Highfield, 2011, Capítulo 13).

Se presenta, entonces, una discontinuidad en el comportamiento de los individuos como miembros de un grupo y como miembros de un grupo que compite con otro. Así pues, la identidad de grupo es un arma de doble filo, porque, por un lado, fomenta la cooperación de los miembros de un grupo al interior de este, pero, por otro, genera actitudes egoístas frente a personas que no hacen parte del grupo. Para evitar esta discontinuidad Dawes & Messick (2000) proponen una formidable solución. Sostienen que al realizar una yuxtaposición de grupos más generales la identidad de grupos puede seguir influyendo en los individuos, en diversas y variadas interacciones sociales. Esto quiere decir que para evitar actitudes racionalmente egoístas entre, por ejemplo, dos individuos pertenecientes a grupos distintos, lo que se debe provocar es que estos se ubiquen en un grupo más abarcante que los anteriores, de tal manera

que puedan compartir identidad de grupo como miembros de uno mismo y se fomente la cooperación. Un grupo más abarcante puede ser la pertenencia a una cultura específica o a la especie humana.

### **1.3.2 Soluciones Estratégicas**

#### 1.3.2.1 Elección de contrapartes

Bajo el marco de las *Soluciones estratégicas* Kollock (1998) propone una que se relaciona con la posibilidad que tienen los individuos de elegir su contraparte o con quién interactuar. Esta simple posibilidad provoca que las personas inmersas en una interacción cooperen más fácilmente, dado que, posiblemente, en la elección de contrapartes los individuos utilicen mecanismos reflexivos que les den alguna garantía de la reciprocidad del otro o mecanismos afectivos que les generen confianza y empatía.

#### 1.3.2.2 Libertad para participar o no en una interacción

Experimentos computacionales como los de Hauert et al (2007) han demostrado que la cooperación puede fomentarse cuando se les da a los participantes en un juego dilemático la posibilidad de retirarse de la interacción y, además, la posibilidad de ingresar a ella cuando lo crean conveniente. De esta forma se puede incrementar la posibilidad de comportamientos guiados por la reciprocidad. También este tipo de estrategias presuponen ciertos tipos de expectativas y preferencias que pueden provocar que una persona decida confiadamente participar en una interacción. Así pues, el hecho de que alguien decida participar en una empresa, puede tener que ver con que tal persona ya tiene ciertas preferencias y expectativas bien definidas.

### **1.3.3 Soluciones estructurales**

#### 1.3.3.1 Identificabilidad

Como vimos anteriormente, la posibilidad del anonimato en interacciones sociales puede fomentar actitudes de *Free rider* y codiciosas, pues existe una muy baja probabilidad de que el individuo explotador pueda ser sancionado mediante castigos o a través de la abstención de la interacción. La identificabilidad de los jugadores permite un incremento en la información que todos tienen de sus comportamientos previos y facilita la creación de interacciones duraderas

con aquellos individuos que son reconocidos como confiables o con buena reputación (Kollock 1998, 199; Nowak & Highfield, 2011, Cap. 2).

#### 1.3.3.2 Manipulación de la capacidad de información

Hemos visto que la falta de información puede provocar que los individuos se abstengan de tomar decisiones cooperativas, porque, por un lado, puede que no sepan con quien están cooperando (Identificabilidad) y, por otro, puede que la falta de información no les deje comprender en qué situación se encuentran y, por ende, qué decisiones tomar para salir de ella. Así pues, la falta de información sobre la situación particular -ya no sobre la contraparte-, sobre la estructura de la situación o sobre las posibilidades de elección e implicaciones de las decisiones, puede disminuir en gran medida el comportamiento cooperativo de los individuos.

Manipulando la cantidad de información relacionada con la estructura de la situación y las posibilidades de elección, algunos estudios (Marwell & Ames, 1979) han mostrado que los temores en la acción o el miedo a ser el tonto (*sucker*) se disminuyen, incrementando de esta manera el comportamiento cooperativo en las interacciones sociales.

#### 1.3.3.3 Eficacia

Kollock (1998) sostiene que si las decisiones de los individuos claramente afectan los resultados obtenidos por el grupo, dichos individuos guiarán su comportamiento a través de una racionalidad colectiva. Esto se debe a que los individuos se sienten más involucrados en la interacción y más impelidos a buscar objetivos comunes, pues comprenden o perciben que sus decisiones tiene un gran efecto en los resultados de la comunidad.

La eficacia puede presentarse fácilmente en grupos o comunidades pequeñas en las que las decisiones de los individuos influyen con evidencia en los resultados de la comunidad.

#### 1.3.3.4 Tamaño del grupo

Diversos estudios han demostrado que hay un incremento de la cooperación cuando los grupos son reducidos, y ocurre un decrecimiento de la misma cuando los grupos son grandes (Rapaport et al, 1962; Bixenstein et al, 1966; Marwell & Schmidt, 1972; Hamburger et al, 1975; Bonacich et al 1976; Fox & Guyer, 1977; Komorita & Lapworth 1982; Dawes, 1980; Kollock, 1998; Nowak & Highfield, 2011). Esto se debe a que en los grupos pequeños la comunicación se facilita entre

miembros, posibilitando los compromisos y la coordinación de preferencias. En los grupos grandes esto se dificulta exponencialmente. Por otro lado, en los grupos pequeños, como ya se dijo, se presenta con facilidad la percepción de eficacia de las decisiones de los individuos, involucrándolos más profundamente en las decisiones mancomunadas. Asimismo, un sistema de vigilancia y monitoreo es menos costoso y mucho más efectivo.

#### 1.3.3.5 Cambio en la estructura de pagos

Como Dawes (1980) lo sostiene, los dilemas sociales son definidos en términos de la estructura de sus pagos, por tanto, la solución más simple será el cambio de dicha estructura. Hay diversas estrategias para realizar un cambio en la estructura de los pagos de una situación dilemática. Aunque Kollock (1998) ubique a las sanciones y recompensas por fuera de esta estrategia, considero que una de las finalidades de las sanciones y las recompensas es modificar la estructura de los pagos en un dilema social. Así pues, un sistema de sanciones o recompensas introduce en el juego dilemático cambios en las preferencias de los jugadores, en unos casos, haciéndolos desistir de inclinaciones egoístas y, en otros, asignando más valor a las decisiones comunitarias.

Por ejemplo, al utilizar incentivos selectivos, definidos por Olson (1965) como bienes privados dados a un individuo para inducir comportamientos cooperativos que generen y soporten bienes públicos, se puede, con un costo determinado, cambiar la estructura de los pagos y, por ende, cambiar también las preferencias de los individuos enfrentados a situaciones dilemáticas.

Así pues, la cooperación es más probable si los individuos tienen la posibilidad de castigar la desertión de los demás. El castigo o la sanción modifican negativamente la estructura de pagos de un dilema, provocando, nuevamente, modificaciones en las preferencias de los individuos.

A pesar de que estas estrategias pueden alcanzar un alto grado de eficiencia en el fomento del comportamiento cooperativo, existe un problema con el costo de instaurar un sistema de sanciones e incentivos (en el capítulo 3 llevo a cabo una propuesta de solución a este problema). Dicho costo deberá ser soportado por la comunidad, disminuyendo los beneficios relativos de la empresa mancomunada. En primer lugar, se presenta un costo directo en la distribución de incentivos y sanciones; en segundo lugar, un sistema de este tipo debe contemplar no solamente el costo directo de las sanciones y los incentivos, sino que también el costo derivado de la vigilancia o monitoreo para sancionar e incentivar eficazmente. Como lo dije, este costo debe restarse a los beneficios obtenidos en la empresa comunal, de tal forma que en ocasiones

puede que los costos hagan que los beneficios esperados por los esfuerzos cooperativos sean tan bajos que es preferible guiarse por intereses particulares (Kollock 1998, 205).

Además, como un problema adicional se encuentra el hecho de que un sistema de sanciones e incentivos se constituye en un bien público de segundo orden, que pretende fomentar la cooperación en bienes públicos de primer orden. Por lo tanto, dado que el sistema es un bien público, este corre el peligro de ser explotado por desertores, de la misma manera que puede ser explotado el bien público de primer orden. Veremos en posteriores capítulos como solucionar este inconveniente.

Un sistema de recompensas y castigos puede proveerse de dos maneras: mediante la vigilancia y coerción mutua entre participantes de una comunidad (Hardin, 1968) o mediante la vigilancia y coerción externa ejercida por un poder o soberano instituido para tal fin (Hobbes, 1651).

Aunque Kollock (1998) sugiere que este tipo de *limitaciones* de las posibilidades de explotación de las personas que conviven en comunidad hacen parte de una categoría distinta de soluciones, denominada por él como “Limitaciones”, considero que es más pertinente ubicar dichos mecanismos bajo la categoría de “Cambio en la estructura de pagos”, pues las limitaciones de este tipo simplemente están dirigidas a instituir un sistema de recompensas y sanciones externas que pretenden modificar la estructura de pagos de una situación dilemática.

Así pues, dichas limitaciones han sido propuestas desde hace mucho como estrategias que pueden romper con los dilemas sociales, modificando la estructura de los pagos en tales situaciones. Y en este sentido, propuestas como la de Hardin (1968), que sostiene que solo la coerción mutua entre miembros de una comunidad puede hacer que los recursos comunes no se depreden, o la de Hobbes (1651), que propone una autoridad externa que ejerce el suficiente poder sobre los integrantes de una comunidad como para hacerlos desistir de su egoísmo racional, son ejemplos relacionados con las *limitaciones* de Kollock y que han sido frecuentemente estudiados como estrategias que pretenden modificar la estructura de pagos en un dilema social.

Con todo, estos dispositivos no solamente cambian la estructura de pagos, también pueden generar coordinación de preferencias, haciendo que los individuos sean más conscientes de que pertenecen a un grupo con ciertas reglas de juego y que si son violadas por ellos terminarían por ser castigados (en el capítulo 3 retomo esta propuesta).

En el último capítulo volveremos especialmente a estas soluciones, pues considero que en ellas podemos encontrar elementos importantes para superar, bajo los presupuestos de la elección racional, los problemas del altruismo en la explicación de la cooperación.

Otra clase de estrategias que pueden caer bajo esta categoría de “cambio en la estructura de pagos” tienen que ver con aquellas que producen sanciones internas en el individuo y que, por tal motivo, modifican sus preferencias en situaciones de dilemas sociales. Bajo esta perspectiva se encuentran las normas sociales. Sin embargo, considero que las normas sociales pueden también hacer parte de las otras dos categorías de soluciones propuestas por Kollock (1998) (soluciones motivacionales y soluciones estrategias). Así que las consideraré especialmente en un apartado distinto.

### **1.3.3.6 Iteración y reciprocidad**

Desde tiempos prehistóricos, siguiendo a Ostrom (1998), la reciprocidad ha surgido como uno de los mecanismos más simples y efectivos para romper con los dilemas sociales. Ya Trivers (1971) demostró que los individuos, incluyendo organismos básicos y animales más complejos, pueden ostentar comportamientos aparentemente altruistas, siempre y cuando, se presenten buenas posibilidades de que exista una retribución en un futuro no muy lejano. Trivers llama a esto *Altruismo Reciproco*. Un modelo más complejo de la reciprocidad es el presentado por Robert Axelrod en su artículo *The evolution of cooperation* de 1984. En este artículo Axelrod postula tres condiciones ambientales indispensables para que pueda abrirse la posibilidad hacia la cooperación mediante la reciprocidad. La primera de ellas es que los individuos deben estar involucrados en una relación en curso; es decir, debe haber múltiples repeticiones de interacción, generando una relación continua entre dos o más individuos. La segunda condición es que los individuos deben ser capaces de identificarse mutuamente. Por último, gracias a la identificación mutua, los individuos deben estar informados sobre el comportamiento de los demás en anteriores interacciones.

Para Dawes (1980) este tipo de comportamientos buscan donar beneficios a los demás, solamente con la condición táctica de obtener recompensas futuras de parte de los otros. Así pues, en la ejecución de un *altruismo táctico*, los individuos pueden buscar el reconocimiento general por sus acciones cooperativas e intentar relacionarse siempre con personas que le aseguren recompensas por su ayuda.

En general, puede reconocerse que a la base de muchos comportamientos cooperativos se encuentra la norma de la reciprocidad (Biel, 2008; Mashima & Takahashi, 2008). Esta norma social, aunque no elimina la tensión entre los intereses racionalmente egoístas de los individuos con los intereses colectivos, si genera cambios estructurales en situaciones dilemáticas que le permite a las personas cooperar condicionalmente.

Una de las condiciones esenciales para que se dé la reciprocidad según Trivers (1971) y Axelrod (1984) tiene que ver con la iteración de las interacciones entre los individuos. En este sentido, puede ubicarse a la reciprocidad, que depende esencialmente de dicha estrategia de iteración, en la categoría de soluciones estructurales, puesto que la iteración modifica la estructura de pagos y de preferencias de los individuos, generando condiciones para que la reciprocidad guíe el comportamiento de cada uno de ellos.

Así pues, aunque Kollock da como ejemplo de solución estratégica, que no modifica la estructura del juego, a la iteración o no de un dilema del prisionero, por el contrario, considero que este tipo de cambios en un dilema repercuten precisamente en la estructura de preferencias de los individuos que se enfrentan a dicha situación; puesto que, en principio, el comportamiento cooperativo emerge en una situación iterada porque los individuos saben que su deserción tiene altas posibilidades de ser *castigada* en un futuro no muy lejano y, también, que su comportamiento cooperativo muy probablemente será, en las siguientes interacciones, *recompensado* con la reciprocidad de la contraparte. Para el caso específico del DP encontramos que dependiendo de la iteración o no de las interacciones se modifica la matriz de preferencias del juego así:

Tabla 1.2 Matriz de preferencias en el DP de una sola jugada.

		Jugador 1	
		Cooperar	Desertar
Jugador 2	Cooperar	2;2	4;1
	Desertar	1;4	3,3

Tabla 1.3. Matriz de preferencias en el dilema del prisionero iterado (DPI).

		Jugador 1	
		Cooperar	Desertar
Jugador 2	Cooperar	4;4	1;3
	Desertar	3;1	2;2

Como se puede apreciar, por causa de la repetición indefinida de las interacciones se efectúa un cambio en la estructura del juego, dado que aparece la posibilidad de *premiar* y *castigar* el comportamiento de los actores. En un juego iterado (Tabla 3), cooperar es la estrategia dominante para los individuos y, por tanto, cooperar mutuamente se convierte en un Equilibrio de Nash. Pero además, cooperar mutuamente es la situación Pareto Eficiente del juego. Así pues, el juego, por la modificación en su estructura, deja de ser dilemático.

#### 1.3.3.7 Generación y refuerzo de normas sociales

Las normas sociales guían el comportamiento de las personas de muy diversas maneras, para, principalmente, regular la vida social de los miembros de un grupo (Coleman, 1990). Así pues, las normas sociales intervienen en las preferencias de un individuo, modificando, en ocasiones, la estructura de pagos o introduciendo un elemento moralizante, humanizador o de reciprocidad en las decisiones particulares.

Como ya hemos visto, comúnmente la norma de la reciprocidad está involucrada en la interacción social (Ostrom, 1998). Esta norma de la reciprocidad sugiere a los individuos un comportamiento francamente cooperativo, pues presupone de los demás una retribución igual o mayor a los beneficios entregados previamente. En este caso las normas sociales podrían ubicarse bajo las *Soluciones Estratégicas* propuestas por Kollock (1998).

Pero, además, las normas sociales en ocasiones toman la forma de códigos de honor que sugieren un fuerte carácter moralizante en aquellos individuos que las asumen (Elster, 1989b). A la vez, como lo afirma Dawes (1980) las normas sociales, en su mayoría, pueden desarrollar castigos en caso de su transgresión. Con todo, las normas tienen la habilidad de *motivar* a las personas en ausencia de cualquier amenaza de censura o sanción. En este caso las normas sociales pueden ubicarse en las *Soluciones Motivacionales* vistas anteriormente.

De la misma manera, vinculadas a las normas sociales se encuentran sanciones internas, como la vergüenza, la indignación, la culpa o la turbación (Elster, 1989b) y también, en muchas ocasiones sanciones externas del tipo ostracismo, señalamientos, etc. Asimismo, el seguimiento de las normas sociales puede acarrear ciertos reconocimientos o recompensas que evidentemente terminan por reportar un beneficio al individuo. El hecho de que a las normas sociales estén vinculados tanto sanciones como recompensas, las ubica en la clase de *Soluciones Estructurales* formuladas por Kollock (1998). Específicamente, las normas sociales, al reforzarse por los castigos, recompensas y emociones, pueden transformar la estructura de situaciones dilemáticas, convirtiendo dichas situaciones en escenarios de coordinación de preferencias e intereses (juego de motivos mixtos), posibilitando, de esta forma, la conducta cooperativa entre individuos.

Por lo anterior, considero que las normas sociales pueden influir, a la vez, en distintos factores que provocan las situaciones dilemáticas de la vida en comunidad. Y en este sentido, se constituyen en un elemento integrador de muchas estrategias de solución a los dilemas sociales.

El siguiente es el esquema resumen de la organización propuesta para las soluciones a dilemas sociales:

### **1.3.4 Soluciones para los Dilemas Sociales**

#### **1. *Soluciones motivacionales***

- **Modificación de las orientaciones del valor social mediante la educación.**
  - o Fortalecimiento del carácter cooperativo y altruista de las personas.
- **Fomento en la comunicación.**
  - o Generación de confianza frente a los otros jugadores.
  - o Fortalecimiento de la moralidad como guía de las acciones.
  - o Acceso a la historia de los demás y sus actuaciones previas.
  - o Humanización de la contraparte.
  - o Permite la discusión de la situación entre los actores.
  - o Facilita la generación de identidad de grupo.

- o Realización de acuerdos y compromisos.
- **Identidad de Grupo**
  - o Compromisos implícitos.

## ***2. Soluciones Estratégicas***

- **Elección de contrapartes**
  - o Mayor seguridad de una retribución (Reciprocidad).
- **Libertar para participar o no en una interacción**
  - o Mayor seguridad de una retribución (Reciprocidad).

## ***3. Soluciones estructurales***

- **Identificabilidad**
  - o Disminución de comportamientos codiciosos o de tipo *free rider*.
- **Manipulación de la capacidad de información.**
  - o Posibilidad de conocimiento de la situación y de sus opciones de solución.
  - o Disminución en el miedo a ser el tonto (*sucker*).
- **Eficacia**
  - o Involucramiento en los objetivos comunes.
- **Tamaño del grupo**
  - o Facilidad en la comunicación.
  - o Disminución del anonimato.
  - o Coordinación de preferencias.
  - o Percepción de eficacia.
  - o Menores costos en sistemas de vigilancia y monitoreo.
- **Cambio en la estructura de pagos**

- o Inclusión de un sistema de premios y castigos.
- o Disminución de comportamientos codiciosos y de tipo *free rider*.
- o Iteración y Reciprocidad.
  - Incremento de la reciprocidad gracias a la posibilidad de recompensar y castigar.
- **Generación y refuerzo de normas sociales**
  - o Introducción de elementos morales que guían la acción.
  - o Motivación para la cooperación en ausencia de censura.
  - o Incremento de la reciprocidad estratégica.
  - o Inclusión de sanciones internas (modificación de estructura de pagos).

## 2. COOPERACIÓN, RECIPROCIDAD Y CASTIGO. UN ENFOQUE EVOLUTIVO

Por décadas, la ortodoxia neoclásica de la economía ha supuesto que el comportamiento humano puede explicarse a partir de la presuposición de que los agentes poseen racionalidad perfecta y acceso a información completa, por lo cual, en toda interacción con los demás, pretenden maximizar sus beneficios. Este supuesto, denominado por los economistas como el *Modelo del Homo Economicus*, sostiene entonces que el único determinante en la acción de las personas es el pago que esperan recibir por dicho comportamiento. Economistas como David Ricardo, Francis Edgeworth y Leon Walras, reconocieron en el egoísmo y la racionalidad perfecta la herramienta más fuerte para explicar el comportamiento humano en el intercambio con los demás. Legado recogido de la postura de Adam Smith en su celebre trabajo *The Wealth of Nations*:

"It is not from the benevolence of the butcher the brewer, or the baker that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity, but to their self-love, and never talk to them of our own necessities, but of their advantages" (*The Wealth of Nations*, 1776, 19).

Pero, ahora bien, si el comportamiento humano ha sido explicado de esta forma por siglos, existen ejemplos recurrentes que nos obligan a preguntarnos por la validez de dichos presupuestos antropológicos. Bajo este marco ¿cómo se explica el comportamiento cooperativo en los seres humanos? Comportamientos ampliamente extendidos en las innumerables culturas y sociedades humanas desde sus inicios, que además es un rasgo a gran escala que nos define como seres humanos, como especie social. La cooperación humana se constituye en una fuerte objeción y evidente anomalía de la lógica imperante en las ciencias económicas que por siglos han interpretado el proceder humano desde una perspectiva eminentemente individualista, en una competencia del más fuerte.

Sorprendentemente, el comportamiento cooperativo humano no solamente es paradójico para la explicación ortodoxa de los economistas, también es una dificultad explicativa dentro de la interpretación clásica de la biología evolutiva. Darwin fundamente su teoría de la evolución en el mecanismo de la *selección natural* que incentiva la competencia entre organismos para la adquisición constante de mejoras adaptativas. Estas mejoras individuales provocan que el

organismo logre mayores posibilidades para su supervivencia y reproducción. La guerra del más fuerte la ganará el que más progenie establezca sobre la faz de la Tierra.

This is de doctrine of Malthus, applied to the whole animal and vegetable kingdoms. As many more individuals of each species are born than can possibly survive: and as, consequently, there is a frequently recurring struggle for existence, it follow that any being, if it vary however slightly in any manner profitable to itself, user the complex and sometime varyings conditions of life, will have a better chance of surviving, and thus be *naturally selected*. (Darwin, 1859, 68).

Tanto los presupuestos del *Homo Economicus* como los de la Selección natural emergen de la lógica individualista de la competencia entre personas u organismos. Lamentablemente, esta lógica se revela como limitada cuando hay que explicar el comportamiento cooperativo de la especie humana que a gran escala y de manera omnipresente se enfrenta a estas dos visiones de profunda influencia en la interpretación de la realidad.

En medio de esta tensión, múltiples experimentos realizados desde la década de los ochenta del siglo pasado han encontrado que los resultados contradicen el supuesto del *Homo Economicus*. Al involucrar a los individuos en juegos como los del Dictador o del *Ultimatum* se ha encontrado que, en su comportamiento, tales personas no siempre buscan el beneficio último o la maximización de su propio interés (Rosas, 2005, 90). Como lo sostienen Ferh & Fischbacher (2005) en experimentos enmarcados en el Dilema del Prisionero (DP) de una sola ronda, con frecuencia la tasa de cooperación se encuentra entre el 40 y 60 %. Más aún, sin necesidad de revisar los experimentos desarrollados por los economistas, con confianza podemos asegurar que en la vida diaria, ya no en el laboratorio, encontramos una gran variedad de situaciones en las que las personas parecen comportarse en contra de su propio interés, generando así grandes dudas con respecto a la cobertura explicativa del supuesto canónico de la economía. En ocasiones las personas colaboran con los demás porque creen injusta una situación, porque consideran que hacen parte de una comunidad o grupo social que debe colaborar mutuamente o porque definitivamente los seres humanos adscribimos a los demás puntos de vista y opiniones igualmente válidas a las propias que merecen toda consideración, entre otras razones.

El supuesto canónico parece implicar una complicación a menudo innecesaria de un intrincado y difícil cálculo racional para explicar el comportamiento de las personas en sociedad. En lugar de esto podemos encontrar situaciones en las que los individuos guían su comportamiento únicamente por preferencias sociales tales como la reciprocidad, la empatía o la aversión a la desigualdad. También, en ocasiones, como sostienen Elster (1991a) y Ostrom (1998), las interacciones de los individuos pueden ser debidas a procesos heurísticos que no alcanzan a involucrarse con la pesadez de un cálculo autointeresado o de presuponer información completa en la situación determinada. Esta última visión es, precisamente, de la que me valgo para expresar mis propuestas en el tercer capítulo. Las anteriores razones pueden esgrimirse para explicar situaciones tales como la votación de los ciudadanos, el pago de los impuestos, el respeto a los acuerdos o contratos incompletos, etc.

Ahora bien, esto no quiere decir que el supuesto canónico de la economía deba desecharse, pues como sostiene Cárdenas (2009) “el *Homo Economicus* ha encontrado un muy limitado soporte empírico, excepto en casos de bienes privados puros y en los cuales el mercado funciona bajo los supuestos de la competencia perfecta” (20), en otras palabras el modelo del hombre eminentemente calculador puede llegar a ser una buena herramienta explicativa en aquellas situaciones de mercado fuertemente competitivo (Ostrom, 1998). Más bien, deberá admitir nuevos modelos de explicación que nos permitan alcanzar una comprensión realista del comportamiento humano en sociedad, sin olvidar que aunque pueda que los individuos tengan un tendencia natural a cooperar, es evidente que no siempre siguen dicha tendencia, con frecuencia colaboran de manera condicional con los demás siguiendo normas de reciprocidad que se enmarcan dentro del conjunto de las preferencias sociales.

En términos generales podemos decir con Herbert Gintis (2000) que el éxito de las relaciones sociales depende, en buena medida, de una rica mezcla de virtudes morales y de intereses materiales, algo que el propio Adam Smith develó en sus obras fundamentales.

Lo anterior sugiere una nueva perceptiva de explicación del comportamiento cooperativo humano a partir de la consideración de una racionalidad acotada (Ostrom, 1998; Elster, 1991b; Axelrod, 1997; Cárdenas, 2009). Desde este punto de vista podemos ver que es más factible llegar a explicaciones exitosas si asumimos la asimetría en la información, la introducción de preferencias sociales (Cárdenas, 2009) y emociones morales (Frank, 1988), que facilitan el empleo de comportamientos heurísticos (Elster, 1991b; Ostrom, 1998) o de tipo ensayo y error

(imitación) (Axelrod, 1997) para asimilar normas cambiantes en la realidad que le permite a las personas interactuar de manera satisfactoriamente benéfica. Esta perspectiva puede fácilmente enmarcarse, para su formalización, en una aproximación evolutiva que además de dar cuenta de dichos comportamientos puede proponer explicaciones de sus orígenes.

## ***2.1 Altruismo parental y selección individual***

Un gran número de teóricos evolutivos han preferido buscar la explicación de los comportamiento altruistas o cooperativos de algunas especies guiándose por la selección individual y dejando a un lado la complicada y poco probable Selección de grupos, que revisaremos posteriormente.

William Hamilton sostuvo que en ciertos casos, si no en la mayoría, los comportamientos altruistas o cooperativos en los animales pueden ser explicados desde la perspectiva darwiniana de la lógica de la selección individual. En su renombrado trabajo de 1964, *The genetical evolution of social behaviour*, Hamilton propone una iluminadora teoría que amplía el concepto darwiniano de “éxito adaptativo” mostrando que los individuos pueden mejorar su aptitud, definida por su tasa reproductiva, ayudando a incrementar la aptitud de parientes cercanos que definitivamente compartan sus mismos genes. A esta estrategia Hamilton la denominó “adaptación inclusiva”. Este tipo de comportamiento puede ser calificado como altruista, en tanto existe un individuo que beneficia a otro, reportando para él mismo un costo directo e inmediato. Sin embargo, indirectamente el éxito reproductivo del actor altruista será efectivo, puesto que la donación de beneficios es efectuada en familiares cercanos que llevan sus mismos genes. De esta forma el comportamiento altruista de tipo parental evoluciona a través de la adaptación inclusiva que favorece la propagación del rasgo comportamental<sup>6</sup> en grupos familiares. Las casta estériles en los himenópteros pueden ser explicadas mediante la teoría de la adaptación inclusiva de Hamilton, pues el particular sistema de apareamiento que reportan, hace que las hembras sean genéticamente más parecidas o cercanas a sus hermanas que a sus madres, facilitando de esta manera que las hembras estériles se dediquen de por vida al cuidado de sus hermanas fértiles.

---

<sup>6</sup> Me atrevo a denominar al altruismo como un rasgo comportamental, apoyado en la equiparación realizada por el propio Hamilton (1963), desde el punto de vista evolutivo, al rededor del comportamiento y los rasgos morfológicos: “se admite por lo general que las características comportamentales de una especie son producto de la evolución en la misma medida que su morfología”.

Hamilton (1963) expresó su teoría en una regla potentemente simple que denominó *Coefficiente de parentesco*. La regla se expresa de la siguiente manera:

$$r \times b > c$$

En donde “r” es la probabilidad que tiene un organismo de poseer un gen que codifique el comportamiento altruista; “b” representa los beneficios brindados a los parientes del individuo que se comporta como altruista y “c” es el costo que se genera para un individuo que se comporta como altruista. Así pues, el costo del altruismo debe compensarse o anularse, por eso el costo “c” aparece como algo menor a los beneficios entregados por el altruista a sus parientes. Dichos beneficios se incrementan a mayor grado de cercanía genética o parentesco. En otras palabras, la *regla de Hamilton* expresa que un gen altruista se propaga en una población si se replica en parientes consanguíneos del portador.

En términos de Trivers (1971) la teoría de Hamilton contempla una fuerte relación genética así:

Hamilton (1964) has demonstrated that degree of relationship is an important parameter in predicting how selection will operate, and behavior which appears altruistic may, on knowledge of the genetic relationships of the organisms involved, be explicable in terms of natural selection: those genes being selected for that contribute to their own perpetuation, regardless of which individual the genes appear in. (Trivers, 35).

La teoría de selección por parentesco (*Kin Selection*) que Hamilton elaboró en su famoso ensayo es explotada hasta la saciedad por Richard Dawkins, quien ha sido uno de los más enconados críticos de la Selección de grupos y defensor, además, del adaptacionismos a nivel inferior.

Dawkins sostiene que: dado que los comportamientos altruistas están regidos por el programa que los genes le imparten al individuo, es decir, que un gen o un grupo de genes son los responsables de codificar el comportamiento altruista en los individuos, entonces, en este sentido, la selección natural trabaja teniendo en cuenta el nivel genético.

En su extenso trabajo, *The Selfish Gen* (1976), Dawkins afirma que un grupo de genes programan al organismo para que su comportamiento se dirija exclusivamente a beneficiar a sus parientes cercanos, en tanto que, existe una alta probabilidad de que los parientes cercanos posean el gen o el grupo de genes altruistas también. Pero de la misma forma, el gen puede

programar al organismo para que beneficie a otros individuos que de alguna manera puedan ser reconocidos como altruistas, esto obedece a que dichos organismos presentan una muy alta probabilidad de tener genes altruistas y, de cualquier modo, se está beneficiando la reproducción de los genes altruistas.

Se muestra entonces que la selección natural al nivel de genes condiciona el comportamiento altruista para que en últimas el beneficiado de dicho comportamiento sea el gen que lo codifica. Por tal motivo, no puede sostenerse, desde este punto de vista, la existencia de un altruismo puro, incondicionado o extremo que de manera indiscriminada done beneficios a cualquier organismo. Esto es algo que el mismo Dawkins sostiene cuando afirma que el gen es eminentemente egoísta, pues busca su beneficio hasta el punto de codificar comportamientos altruistas para su supervivencia.

## ***2.2 Altruismo Recíproco***

Dawkins (1976) sostiene que los genes pueden codificar comportamiento que reporten beneficios a individuos que están más allá del círculo familiar del actor, algo que Hamilton no había considerado, pero que gracias a su influencia, Trivers en 1971 saca a la luz. En *The Evolution of Reciprocal Altruism* (1971) elabora una teoría que pretende extender los alcances del altruismo parental hacia un altruismo recíproco.

El modelo de Trivers pretende salirse de los límites del parentesco para explicar comportamientos altruistas que a pesar de la teoría de Hamilton se mantuvieron en la oscuridad. Es el caso, por ejemplo, de la llamada de alarma de ciertas aves, el comportamiento involucrado en simbiosis de limpieza entre distintas especies y, por supuesto, el extendido comportamiento cooperativo en humanos.

The model presented here is designed to show how certain classes of behavior conveniently denoted as altruistic (or reciprocally altruistic") can be selected for even when the recipient is so distantly related to the organism performing the altruistic act that kin selection can be ruled out. (Trivers, 35).

Es bueno anotar que el altruismo al que se refiere Trivers (1971) es evidentemente limitado o menos puro, pues la cooperación de los individuos está supeditada a la amplia posibilidad de que exista una retribución en un futuro no muy lejano.

Trivers (1971) sostiene que hay tres posibilidades o tipos de comportamiento altruista: a) cuando el altruista dispersa los beneficios aleatoriamente por toda la población, (altruismo puro o incondicional), b) cuando el altruista dispersa los beneficios no aleatoriamente, considerando el grado de relación genética con los posibles receptores, (altruismo parental) y c) cuando el altruista dispersa los beneficios no aleatoriamente, considerando las tendencias altruistas de los posibles beneficiarios o receptores, (altruismo recíproco). (Trivers, 1971, 36).

Ahora bien, Trivers afirma que para que se dé el último caso deben cumplirse tres condiciones, definidas por él así:

1. Deben haber numerosas situaciones de tipo altruista en la expectativa de vida del altruista.
2. Un determinado altruista debe actuar repetidamente con el mismo conjunto de individuos.
3. Pares de altruistas deben estar expuestos "simétricamente" a situaciones altruistas, para que ambos presten de forma aproximada beneficios equivalentes uno al otro y sufran costos también equivalentes.

Si estas tres condiciones se presentan, sostiene Trivers (1971), el comportamiento altruista del tipo recíproco puede ser seleccionado. Dichas condiciones se ven cumplidas en la naturaleza como lo demuestra Triver así:

La primera condición puede verse representada en la expectativa de vida larga de ciertos organismos. Dada tal expectativa se maximiza la posibilidad de que dos individuos puedan interactuar más de una vez, y que esta interacción sea una situación altruista que puede repetirse, generando la posibilidad de la reciprocidad en el futuro.

La segunda condición puede cumplirse si observamos la baja tasa de dispersión de ciertas especies que posibilita que un individuo interactúe repetidamente con el mismo conjunto pequeño de individuos vecinos, provocando que el comportamiento recíproco sea factible.

Dawkins denomina a este fenómeno "viscosidad", explicando que es la tendencia de los individuos para continuar viviendo en el lugar de su nacimiento (Dawkins, 1989).

La tercera condición es un poco más compleja y, según Trivers, podemos encontrarla en casos típicos de interacción no-jerárquica en donde no hay relaciones de poder involucradas o, por ejemplo, en casos en donde la jerarquía si se presenta, pero los que ostentan el poder pueden verse necesitados de la ayuda de los demás, y los que no lo ostentan pueden verse beneficiados de la protección del poderoso. La ayuda en combate resulta ser para Trivers un buen ejemplo de simetría en relaciones jerárquicas (Trivers, 38).

El modelo de Trivers predice que el beneficio de la reciprocidad en un futuro será mucho mayor que el costo de la colaboración inicial y, evidentemente, el costo de la no colaboración será mucho mayor que el beneficio de mantener un comportamiento egoísta. Según Trivers, gracias a la ventaja proporcional del comportamiento altruista recíproco con respecto a la del comportamiento egoísta, y a la posibilidad de que dicho comportamiento recíproco pueda seguir dándose entre dos o más actores, la presión selectiva de la naturaleza favorecerá a aquellos individuos que sean capaces de cooperar discriminadamente con individuos que a su vez también cooperan.

Ahora bien, la clave del éxito del altruismo recíproco se fundamenta en la posibilidad de ejercer una interacción discriminada entre individuos, que permita reconocer con certeza a los egoístas y privilegiar las interacciones con los altruistas. Precisamente, las tres condiciones del modelo de Trivers pretenden delimitar en un primer momento la interacción discriminada.

Para el caso del comportamiento cooperativo en la especie humana, tanto Trivers (1971), Alexander (1987) como Sober & Wilson (1998), están de acuerdo con que este tipo de interacción involucra mucho más que las tres condiciones expuestas en el modelo del altruismo recíproco. Involucran "cierto nivel de sofisticación cognitiva para percibir las tendencias altruistas de los demás" (Sober & Wilson, 1998, 115) y las tendencias explotadoras o tramposas, también.

Según afirma Trivers (1971), la selección favorecerá comportamiento que puedan explotar los beneficios del altruismo siempre y cuando dicha actitud pueda quedar impune. Es decir, que si un organismo es capaz de interactuar con un altruista, recibiendo los beneficios que le otorga

esta interacción, pero sin dar nada a cambio o, por lo menos, nada equitativo, dicho comportamiento será más apto que el del altruista mismo. Empero, a su vez, sostiene Trivers (1971), si el altruista genera un mecanismo que le permita identificar con precisión a este tipo de explotadores o tramposos la selección favorecerá este rasgo, puesto que dicho mecanismo incrementa el *fitness* del altruista y disminuye el del tramposo. En conclusión, así como la selección individual favorece mecanismos más elaborados para explotar la interacción (hacer trampa), también favorece mecanismos, igualmente elaborados, que contrarrestan esta explotación y que permiten identificar al tramposo, permitiendo de forma confiable una interacción discriminada.

Mecanismos cognitivos como la memoria o la capacidad de distinguir rostros y prevenirse ante la apariencia de un individuo, son formas que los seres humanos, y tal vez algunas especies de simios mayores, emplean para acceder a una interacción discriminada y a una cooperación más estable.

Pero, además de este tipo de mecanismos cognitivos, Trivers (1971) sostiene que gracias a la evolución de la cooperación humana, aparecen sentimientos morales como la amistad, la indignación, la culpa o la agresión moral que refuerzan el comportamiento cooperativo. Pues, por ejemplo, en el caso de que un individuo sea identificado o reconocido como tramposo en una situación dada, la culpa y el arrepentimiento que este individuo siente al no cooperar, o mejor dicho, al haber sido identificado como tramposo, le permitirá tener posibilidades nuevas de interacción con individuos altruistas que, gracias a sus sentimientos de amistad, aprecio y perdón al reconocer el arrepentimiento del otro, retomarán la interacción en vista de los beneficios futuros que ésta reporta a ambas partes.

### ***2.3 El Dilema del Prisionero Iterado (DIP)***

En un artículo conjunto, Robert Axelrod & William Hamilton (1981) retoman la propuesta presentada por Trivers en 1971, realizando un estudio más profundo desde la perspectiva de la teoría de juegos (que el propio Trivers ya había utilizado en su famoso artículo), y proponiendo que el altruismo recíproco puede ser ejecutado por organismos que no poseen capacidades cognitivas complejas o sofisticadas, como es el caso de virus y en general poblaciones microbianas.

Axelrod & Hamilton inician su artículo *The Evolution of Cooperation* (1981) refiriéndose al error que se ha cometido al considerar a la cooperación como una adaptación a nivel de grupos, es decir, un rasgo favorecido por un proceso de selección a nivel superior. Estos dos autores sostienen que "Recent reviews of the evolutionary process, however, have shown no sound basis for a pervasive group benefit view of selection; at the level of a species or a population, the processes of selection are weak" (Axelrod & Hamilton, 1981, 1390).

El objetivo primordial de Axelrod & Hamilton (1981), empero, tiene que ver con alcanzar una comprensión más rica de la evolución de la reciprocidad, revisando a partir de la teoría de juegos, particularmente del DP, la estabilidad final de una estrategia cooperativa, su robustez y su viabilidad.

Razonan los autores que siempre y cuando las interacciones que se dan en la naturaleza sean aleatorias, es decir, con una muy pequeña posibilidad de que las interacciones se repitan entre los mismos actores, es de suponer que en un DP la estrategia más estable ha de ser siempre desertar, pues no hay incentivos para la cooperación, en tanto la posibilidad de la reciprocidad es mínima.

Apart from being the solution in game theory, defection is also the solution in biological evolution. It is the outcome of inevitable evolutionary trends through mutation and natural selection: if the payoffs are in terms of fitness, and the interactions between pairs of individuals are random and not repeated, then any population with a mixture of heritable strategies evolves to a state where all individuals are defectors. [...] In these respects the strategy of defection is stable. (Axelrod & Hamilton, 1981, 1392).

Ahora bien, Axelrod & Hamilton (1981) resaltan que existen muchas situaciones biológicas en las que las interacciones llevadas a cabo por la mayoría de los organismos tienen la forma del DP, pero con una particularidad: en tales casos las interacciones se pueden dar más de una vez y repetidamente, algo que Trivers (1971), en las dos primeras condiciones para su modelo de altruismo recíproco, ya había puesto de manifiesto. Axelrod & Hamilton (1981) utilizan, entonces, una nueva versión del DP que se acomoda más a estas situaciones, denominada Dilema del Prisionero Iterado (DPI). Este nuevo modelo tiene en cuenta lo que Axelrod designa como la "sombra del futuro", es decir, la posibilidad de realizar un número indeterminado de repetidas interacciones (jugadas) entre los mismos actores. En principio, estas situaciones

biológicas, evidentemente son posibles, siempre y cuando los individuos que interactúan posean una capacidad cognitiva algo elaborada que les permita recordar interacciones anteriores y reconocer a individuos con quienes previamente han interactuado, algo sobre lo que Trivers (1971) y Sober & Wilson (1998) han llamado la atención. Si estas capacidades cognitivas no están presentes en los individuos que interactúan, por más que dichas interacciones se den repetidamente, para cada uno de ellos surgirán como una nueva interacción sin historial ni futuro.

En contra de esta posición Axelrod & Hamilton (1981) sostienen que no es necesario que los organismos tengan cerebro o una capacidad de previsión para que puedan utilizar estrategias de interacción diferencial en la realidad biológica. Las bacterias son un ejemplo de esto:

(i) Bacteria are highly responsive to selected aspects of their environment; (ii) this implies that they can respond differentially to what other organisms around them are doing; (iii) these conditional strategies of behavior can certainly be inherited; and (iv) the behavior of a bacterium can affect the fitness of other organisms around it, just as the behavior of other organisms can affect the fitness of a bacterium. (Axelrod & Hamilton 1981, 1392)

Desde este punto de vista, el DPI ha de contemplar, en una forma mucho menos restrictiva, interacciones entre bacterias, por ejemplo, o entre una colonia de bacterias y primates, en tanto que la condición de la previsión para una interacción discriminada puede ser remplazada por estrategias diversas que organismos sin esta capacidad pueden emplear.

Para que la cooperación surja han de presentarse tres rasgos importantes, y en cierto modo imprescindibles, para su selección y permanencia en la naturaleza. Axelrod & Hamilton (1981) afirman, como ya lo mencionamos anteriormente, que la estrategia de la cooperación debe ser viable, robusta y evolutivamente estable.

La viabilidad se refiere a la posibilidad de que una estrategia, en este caso la cooperativa, pueda encontrar oportunidad de desempeñarse en medio de un ambiente completamente adverso, por ejemplo, un ambiente de egoístas. La segunda característica tiene que ver con que la estrategia de cooperación debe prosperar en medio de un ambiente altamente diversificado de estrategias que compiten con ella. La última característica enunciada, fue definida en 1982 por el biólogo inglés Maynard Smith quien, utilizando la teoría de juegos, postuló que una Estrategia

Evolutivamente Estable (EEE) es aquella que es adoptada por la mayoría de integrantes de una población y no puede ser invadida o mejorada por ninguna estrategia mutante o alternativa.

Axelrod organizó una competencia computacional entre diversas estrategias para determinar cuál era la más eficiente en el DPI. Resultó que, de las 15 estrategias propuestas por expertos en teoría de juegos, la denominada *Tit for Tat*, elaborada por el psicólogo ruso Anatol Rapoport, fue la ganadora. Esta estrategia se caracteriza por iniciar la cooperación en la primera interacción y copiar el comportamiento previo de su adversario permanentemente. Así pues, cuando el adversario no coopera, la estrategia le indicará al jugador que no coopere en la siguiente jugada y si el adversario coopera, esto es garantía de que el jugador también cooperará. Este forma de "ojo por ojo" permite que la represalia de un jugador establezca la interacción, en tanto que el adversario no podrá aprovecharse inequitativamente de su contraparte, y, más bien, lo empujará a cooperar para recibir los beneficios de una interacción recíproca cooperativa prolongada.

Para Axelrod & Hamilton (1981) no solamente *Tit for Tat* es eficiente frente a las demás estrategias, sino que es la mejor por ser viable, robusta y Evolutivamente Estable.

El hecho de que haya triunfado frente a las demás estrategias, que por cierto en un segundo torneo fueron 62 y que también derrotó, demuestra su robustez, pues prosperó en un ambiente altamente diversificado de estrategias rivales. Su éxito, como lo sostiene el autor, radica en que es una estrategia eminentemente "amable", pues inicia la cooperación y la mantiene a menos que su contraparte no coopere, en cuyo caso toma una posición de retaliación inmediata, para posteriormente cooperar si su oponente así lo hace. Este carácter de "amabilidad" hace que *Tit for Tat* evite el conflicto o, si llega a darse este, evita que se prolongue innecesariamente, pues nunca lo provoca al no desertar, a menos que el otro lo haya hecho previamente. Por esto mismo es una estrategia transparente y permite que el otro jugador actúe basado en la confianza de la interacción. También tiene la capacidad de "olvidar" la no cooperación si la contraparte decide volver a cooperar.

Si además de lo anterior se añade la "sombra del futuro", es decir la inmensa posibilidad de que existan muchas interacciones en el futuro entre los mismos individuos, se puede colegir que la estrategia *Tit for Tat* es Evolutivamente Estable. La "sombra del futuro" les permite a los actores mantener una actitud cooperativa, en tanto, existe la confianza de que las ventajas de la

cooperación puedan percibirse por tiempo indefinido. Si dicha interacción tiene un límite de repetición conocido por los actores, estos emplearán la estrategia de siempre desertar o no cooperar dada la seguridad de que la reciprocidad tiene un término definido y, por tanto, se querrá sacar más provecho desertando antes de llegar a ese límite.

Sin embargo, como lo sostienen Axelrod & Hamilton (1981), *Tit for Tat* no es la única EEE, pues SIEMPRE D (siempre no desertar) también lo es, sin importar la posibilidad de que continúe o no la interacción. En el DP de una sola ronda está demostrado que no cooperar es la estrategia más racional, de la misma manera, en una versión iterada del Dilema SIEMPRE D es evolutivamente estable (Boyd, et al, 2005; Fehr & Fischbacher, 2005), pues puede sacar gran provecho de la incauta cooperación de otros y, por otro lado, aunque tal cooperación no se presente, el jugador que utilice SIEMPRE D nunca perderá.

¿Cómo es posible que *Tit for Tat* pueda surgir en medio de un ambiente dominado por SIEMPRE D? La viabilidad de *Tit for Tat* se ve comprometida en tanto que SIEMPRE D es evolutivamente estable y ampliamente extendida en la naturaleza. Para Axelrod & Hamilton (1981) existe una salida al equilibrio generado por SIEMPRE D, dicha salida es la cooperación surgida en pequeños grupos de individuos emparentados. La teoría del parentesco genético podría ser una respuesta a la viabilidad de la cooperación en el mundo natural, dado que dicha estrategia pudo generarse y contraponerse a la fuerza de SIEMPRE D gracias a los lazos genéticos entre pequeños grupos familiares. Una vez aparecida en estos grupos la estrategia cooperativa, sus ventajas adaptativas pudieron extenderse a grupos cada vez más numerosos y con menos individuos relacionados entre sí que, sin embargo, tenían una probabilidad alta de reencontrarse e interactuar. En otras palabras, SIEMPRE D puede ser invadida por *Tit for Tat* gracias al parentesco genético de un pequeño grupo y, posteriormente, gracias a la alta probabilidad de que las interacciones entre dos individuos puedan ser repetidas, para esto es necesario que exista una especie de "apiñamiento" de la población o, como lo propone Dawkins, se debe presentar viscosidad poblacional para que los integrantes de dicho grupo tengan una probabilidad no trivial de volver a interactuar.

La reciprocidad o altruismo recíproco así alcanzado puede caracterizarse por cuatro elementos, a saber: a) presencia de sombra del futuro, posibilidad de futuras interacciones, b) probabilidad de que los individuos con los que tal vez se interactuará sean cooperadores condicionales, c) la

decisión a cooperar inicialmente con el otro y d) rechazo a cooperar si el otro en interacciones previas ha defraudado los actos cooperativos con la deserción.

Fehr, Fischbacher & Gächter (2002) describen al altruismo recíproco como aquel en donde un actor condiciona su comportamiento sobre el conocimiento del comportamiento previo de los otros. En este sentido, el actor altruista recíproco, a diferencia del altruista incondicional, solo ofrecerá su ayuda y colaboración a otros individuos, inclusive a aquellos no emparentados con él, si espera algún tipo de beneficio en el futuro.

Hay que anotar, por lo demás, que el modelo descrito por Axelrod (1984) tiene como principal característica el hecho de que se desarrolla en interacciones bipersonales, tal como está diseñado el clásico DP. Por otro lado, y como una extensión de lo hecho por Axelrod & Hamilton (1981), múltiples estudios (Joshi, 1987; Bendor & Mookherjee, 1987; Boyd & Richerson, 1989, 1992; Fehr & Fishbacher, 2005; Ostrom, 1998; Hirshleifer & Rasmusen, 1989; Yamahashi & Takajashi, 1994) han concluido que las condiciones en las que el altruismo recíproco o la reciprocidad puede evolucionar se relacionan, a lo sumo, con la presencia de grupos pequeños en donde la viscosidad poblacional es alta.

## ***2.4 Problemas para la estabilidad de la reciprocidad***

Hemos visto cómo se han generado diversas propuestas, desde las ciencias biológicas, pasando por la economía hasta la teoría de juegos, de modelos de explicación de la evolución de la cooperación. Sin embargo, gracias a evidencias experimentales y de la vida real se ha encontrado que cada uno de estos modelos tiene sus propias limitaciones para explicar el comportamiento cooperativo en los seres humanos.

Asimismo, como ya se mencionó, todos los modelos están basados en interacciones bipersonales que no reflejan de forma adecuada la vida social en grupos en donde las relaciones se entretajan a través de múltiples interacciones simultáneas. Aunque es correcto afirmar que en interacciones bilaterales repetidas la reciprocidad puede llegar a ser una estrategia evolutivamente estable, también es cierto que cuando se pasa al nivel de interacciones N-personales dicha estrategia decae frente a la aparición de oportunistas que desertan aprovechándose de los beneficios donados por los reciprocadores (Boyd, et al, 2005; Fehr &

Fischbacher, 2005), estos oportunistas han sido denominados a través de la vasta literatura sobre el tema como *Free riders*.

Una posible explicación de este fenómeno tiene que ver con que en grupos grandes el anonimato de sus miembros es más probable, dando lugar a que los incentivos materiales para la desertión dominen el comportamiento de algunos individuos al poder actuar de forma injusta o en contra del bien público, explotando la colaboración de los demás sin que sean detectados o con una muy baja probabilidad para esto. Los incentivos para el oportunista se potencializan cuando se presenta una situación en donde es imposible o muy difícil limitar el acceso al bien común o público (Fehr & Fischbacher, 2005) y, además, existe la configuración contextual que permite cierto grado de anonimato en las acciones de los individuos que intervienen en la provisión, mantenimiento o aprovechamiento del bien.

Dado que como lo afirman Fehr, Fischbacher & Gächter (2002), en las sociedades humanas parece ser la regla y no la excepción que se presenten permanentemente incentivos materiales para desertar, una estrategia como la reciprocidad está enfrentada a constantes amenazas de los *Free riders*, hasta el punto de que a lo largo del tiempo dicha estrategia decaerá y cederá su posición a la desertión o explotación.

Lo anterior ha sido demostrado en abundantes experimentos que evidencian que a largo plazo la estrategia del altruismo recíproco en juegos de bienes públicos o comunales decae inevitablemente (Boyd, 1988; Fehr & Fischbacher, 2005). Esto se debe, principalmente, a que durante las repetidas interacciones las personas van adquiriendo más información de sus contrapartes, dejando al descubierto posibles desertiones pasadas que, dado el comportamiento recíproco, van minando la estabilidad en la cooperación hasta el punto de decaer por completo bajo la presión de los explotadores u oportunistas.

### ***2.5 Castigo como solución a la inestabilidad de la cooperación por reciprocidad***

Como lo sostienen Boyd et. al. (2005), se ha encontrado un mecanismo evolutivo que puede dar solución al rompecabezas configurado por la reciprocidad, el oportunismo, el tamaño de los grupos sociales, el anonimato incitado por este, y la débil sombra de futuro que en ocasiones se presenta en la interacción humana. Dicho mecanismo es denominado por los investigadores

como *Castigo Altruista*; definido como la disposición a castigar a aquellos que explotan la cooperación de los demás, forzándolos a cooperar para evitar futuros castigos, incluso si dicho castigo representa un determinado costo para el que lo ejecuta, de tal manera que, frente a individuos que no castigan y que no son castigados, aquellos que si castigan poseen una desventaja adaptativa -precisamente este es el carácter altruista del castigo- (Fehr, Fischbacher & Gächter, 2002; Fehr & Fischbacher, 2005, Boyd, 1988; Boyd & Richerson, 1992; Hauert et al, 2007; Gintis, 2008; Gintis et al, 2005). Con todo, investigadores como Hernich et al, (2006) o Boehm (1993) han encontrado pruebas relacionadas con que el castigo es una practica extendida en múltiples y diversas sociedades. Los castigos pueden ser desde señalamientos, multas, daños físicos, suspensión de beneficios sociales, hasta el ostracismo a personas que comenten delitos o faltas contra el grupo social (Boehm, 1993; Bowles & Gintis, 2003; Fehr & Fischbacher, 2005). Un ejemplo de señalamiento y ostracismo es el referido por Fehr & Fischbacher (2005):

During World War I, British men who did not volunteer for the army faced strong public contempt and were called “whimps” (Fehr y Fischbacher, 2005, 171).

Se ha encontrado que comportamientos de este orden pueden incrementar la tasa de cooperación en grupos en donde inicialmente hay una frecuencia alta de no-cooperadores (Boyd & Richerson, 1992). Esto se debe principalmente a que las ganancias de los *Free riders* obtenidas por la explotación a la cooperación de los demás se ve altamente disminuida al recibir un castigo por su comportamiento, dejándolos en desventaja adaptativa frente a aquellos que cooperan mutuamente. De esta forma, dado que los castigadores también incurren en costos para implementar el castigo a los *Free riders*, la estrategia con mejores resultados será la de cooperar y abstenerse de castigar.

Supongamos que tenemos una población de individuos en donde se presentan las tres estrategias, a saber, Cooperadores (C), *Free riders* (F) y Castigadores (P). Los (C) obtendrán un beneficio  $b = 3$  por cada vez que interactúen con otros (C) o con otros (P) -no olvidemos que estos últimos también son cooperadores-, pero cuando interactúen con (F) estos, los *Free riders*, obtendrán un beneficio de explotación  $b_e = 6$  y los (C) no obtendrán ningún beneficio por haber sido explotado  $e = 0$ . De la misma forma los (P) que interactúen con otros (P) o con otros (C) obtendrán un  $b = 3$ , pero cuando interactúen con (F) no obtendrán ningún beneficio, por el contrario incurrirán en un costo para infringir castigo a los (F) de  $c_p = -2$ , mientras que a los (F) se les infringirá un castigo que representa un costo  $c = -4$ . Por último, si dos (F) interactúan, ambos

se explotaran y, por lo tanto, no obtendrán ningún beneficio por haberse explotado mutuamente  $e = 0$ .

Si todos interactúan con todos, podemos hallar los siguientes resultados:

Tabla 2.1

Resultados en interacción indiscriminada de Cooperadores (C), <i>Free riders</i> (F) y Castigadores (P)
R(C) $b + e + b = 6$
R(F) $b_e + e + c = 2$
R(P) $b + c_p + b = 4$

Teniendo en cuenta esta situación hipotética que resume, en buena medida, los modelos elaborados por los investigadores para demostrar el poder del castigo altruista en el incremento de las tasas de cooperación (Axelrod, 2007; Fehr, Fischbacher & Gächter, 2002; Fehr & Fischbacher, 2005; Fehr & Gächter, 2002; Boyd & Richerson 1992; Boyd et al, 2005; Gintis, 2008; Rosas, 2008; Sripada, 2005), podemos decir que la estrategia que obtiene mejores resultados de una interacción aleatoria es la (C), puesto que aunque puede llegar a ser explotada, por lo menos, no incurre en costos, dándole a sus resultados valores muy por encima de los otros. Al introducir el castigo altruista, una población con esta configuración puede derivar, con interacciones repetidas, en una composición alta de cooperadores y baja de desertores o *Free riders*. La cooperación, entonces, puede ser fomentada por la aparición de castigo altruista (Boyd & Richerson, 1992; Boyd, et al 2005; Henrich & Boyd, 2001; Rosas, 2007; Gintis et al, 2005). Desde otra perspectiva, si es castigado, el *Free rider* tiene incentivos para dejar la explotación, de tal suerte que su mejor elección será evitar el castigo mediante un comportamiento cooperativo. En conclusión, el castigo puede ser un mecanismo que refuerza el seguimiento de la norma de reciprocidad incrementando así la tasa de cooperación en un grupo social.

Simulaciones de juegos de bienes públicos (DP de N-Personas) realizadas por computador han encontrado que la cooperación declina a medida que transcurre el tiempo, pero que al introducir mecanismos de castigo a los desertores la tasa de cooperación se mantiene (Fehr & Gächter, 2002). Ver Figura 2.2.

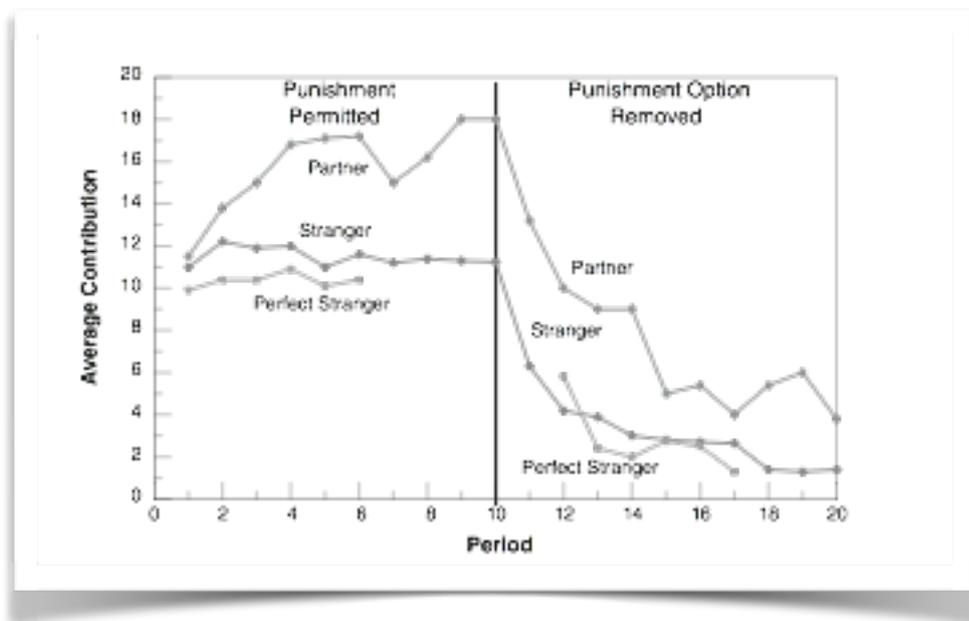


Figura 2.2. La

simulación realizada por Fehr & Gächter (2002) muestra cómo en el tiempo, la tasa de contribución decae cuando no hay opción de castigar a los desertores (lado derecho de la gráfica) y cómo se mantiene cuando el castigo a los desertores es posible (lado izquierdo de la gráfica).

En la actualidad, para la gran mayoría de investigadores interesados en el tema, el comportamiento caracterizado por estar dispuesto a cooperar en interacciones sociales y a castigar la no cooperación, incluso si es necesario a costa de su propio beneficio es conocido como *Reciprocidad Fuerte* (Gintis et al, 2005; Falk & Fischbacher, 2006; Sethi & Somanathan 1996; Bowles & Gintis 2003; Fehr & Henrich 2003; Fehr, Fischbacher & Gächter 2002).

Pero, el castigo altruista involucrado en la *Reciprocidad Fuerte* de los individuos se convierte en un nuevo rompecabezas, pues ¿cómo es posible que ciertos individuos se involucren en comportamientos de castigo a no-cooperadores, a pesar de que dichos comportamientos les reporte costos que los pueden poner en desventaja con otras personas? Algunos autores han intentado responder a este cuestionamiento defendiendo un mecanismo de la Selección natural,

que aún hoy en día es algo controvertido, a saber, la Selección de grupos (Boyd & Richerson 1992; Gintis et al, 2005; Gintis 2008; Sethi y Somanathan, 1996; Fehr & Gächter, 2002; Fehr & Fischbacher, 2005).

En este sentido, la selección de grupos se esgrime como razón por la cual un comportamiento evidentemente maladaptativo para un individuo, como el de incurrir en costos para castigar la deserción de otro y, a su vez, beneficiar a los integrantes de su grupo, es viable en la naturaleza, específicamente en la interacción humana. Boyd & Richerson (1992) argumentan que los grupos con gran número de castigadores presentan altas tasas de comportamiento cooperativo representando esta situación ventajas frente a otros grupos que tiene bajos números de castigadores. En este sentido la frecuencia del castigo y el comportamiento cooperativo es positivamente correlacionado a través de los grupos. En otras palabras, el castigo, como una “respuesta correlacionada” con la cooperación, favorece a los grupos, pues a mayor cooperación mayor *fitness* de grupo (217). De esta manera, los grupo con mayor cantidad de cooperadores prevalecerán en la competencia natural frente a aquellos grupos que tiene una baja tasa de cooperación. Todo esto se debe, según los investigadores, a que la selección individual al interior de los grupos es débil frente a la fuerza selectiva de la competencia entre grupos, provocando que el comportamiento castigador pueda llegar a ser común dentro de los grupos.

En el capítulo tres intento ofrecer argumentos en favor de la estabilidad del castigo sin acudir a la explicación de Selección de grupos. Por el contrario, bajo los presupuestos metodológicos de la racionalidad acotada presento una interpretación adecuada de las sanciones e incentivos como dispositivos que fomentan la cooperación en los seres humanos.

## ***2.6 Dilemas de segundo orden en juegos de bienes públicos y comunales***

Ahora bien, si revisamos con cuidado los resultados de la tabla 2.1, podemos ver que los valores del comportamiento castigador (P) no son los mejores, dado que debe incurrir en costos para castigar a los *Free riders* (F). Esta situación hace que en términos evolutivos, la mejor estrategia sea la de cooperar (C). Desde este punto de vista, no solo los *Free riders* castigados tendrán un incentivo negativo para comportarse de forma cooperadora, sino que también los castigadores (P) tendrán incentivos para dejar de castigar y simplemente dedicarse a cooperar, pues los resultados de los cooperadores no-castigadores son mejores que los demás. Así pues, la

institución del castigo, que incrementa la cooperación en grupos con desertores, se convierte en un bien público de segundo orden (Rosas 2007; Boyd & Richerson 1992; Sober & Wilson 1998), donde el comportamiento que evita castigar por no incurrir en costos puede explotar el comportamiento castigador, generando de esta forma un dilema social de segundo orden (Oliver, 1980; Yamagishi, 1986; Axelrod 1986). Esto provocaría que se disminuyeran los castigadores, incrementando correlativamente la tasa de desertores y, consecuentemente, haciendo decaer la cooperación dentro del grupo social y, como lo dice Axelrod (1986), destruyendo toda moderación que se hubiera generado al inicio de la interacción.

Estudios experimentales (Axelrod, 1986) han demostrado que el castigo altruista, aunque puede mantener la cooperación en largos periodos de tiempo, no es capaz de hacer que esta se convierta en un comportamiento evolutivamente estable, dado que a medida que los castigadores infringen castigo a los no-cooperadores, estos últimos van disminuyendo su frecuencia en la población, provocando, no solamente un incremento en la tasa de cooperación, sino que también una disminución de la población de los propios castigadores dada la escasez de *Free riders*. Al disminuir la población de castigadores a niveles críticos los incentivos para desertar y explotar la cooperación vuelven a ser importantes y nuevamente la tasa de *Free rider* puede incrementarse.

La simulación realizada por Axelrod (1986) fue desarrollada de tal manera que grupos de jugadores interactuaban en dos periodos distintos bajo el marco de un juego de bienes públicos. En el primer periodo los individuos tenían la posibilidad palmaria de cooperar o no en la interacción con otros individuos y en el segundo, los individuos que habían cooperado en el primer periodo tenían la posibilidad de castigar, incurriendo en un costo, a los individuos que no habían cooperado en la interacción previa. Lo que encontró Axelrod (1986) en esta interesante simulación fue que, como otros estudios lo han demostrado (Axelrod, 2007; Fehr, Fischbacher & Gächter 2002; Fehr & Fischbacher 2004 ; 2005; Fehr & Gächter 2002; Boyd, 1988; Boyd & Richerson, 1992; Boyd et al, 2005; Gintis, 2008; Rosas, 2008; Sripada, 2005), el castigo puede incrementar las tasas de cooperación en donde existen desertores (comportamiento audaz como lo denominó Axelrod en su artículo), pero además encontró que después de cien interacciones la estrategia de no castigar y solamente cooperar explota y disminuye la posibilidad de cooperar. Se puede colegir, con los datos de Axelrod (1986), que ser vengativo, castigar en nuestros términos, tampoco es una estrategia evolutivamente estable, pues está

expuesta, como cualquier bien público o comunal, a la explotación de los *Free riders* en un nivel superior.

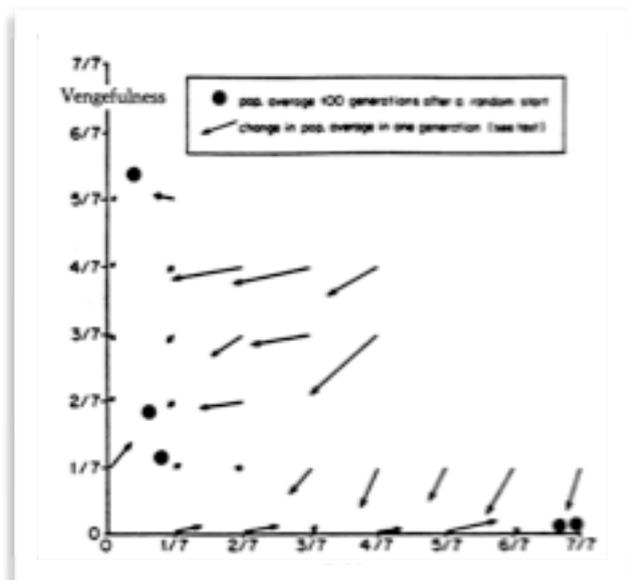


Figura 2.3. La figura muestra cómo el castigo

altruista a los no-cooperadores no puede mantener el incremento en la cooperación en largos periodos de tiempo. Las flechas dirigidas del centro de la gráfica hacia los ejes indican que el grado de vengatividad disminuye a lo largo del tiempo y, correlativamente, el grado de audacia aumenta nuevamente. (Axelrod 1986, 1100).

El propio Axelrod propone una solución a este dilema de segundo orden. En el mismo artículo de 1986, el autor sostiene que una “*metanorma*” puede hacer que el sistema de castigo alcance la auto-vigilancia, de tal suerte que el comportamiento castigador no decaiga en el transcurso del tiempo debido a la deserción de los castigadores hacia la mera cooperación. La *metanorma* definida por Axelrod (1986) no es otra cosa que un comportamiento que no solamente está dispuesto a castigar a los no-cooperadores, sino que además está dispuesto a castigar a los que no castiga la deserción. En otras palabras, la *metanorma* de Axelrod (1986) es representada por un comportamiento que también castiga a los no-castigadores. Boyd & Richerson (1992) lo llamaran con posterioridad: comportamiento de “*retribución*” o “*estrategia moralista*”. Este nuevo mecanismo se enmarca también en el denominado comportamiento de *Reciprocidad Fuerte* de los seres humanos.

Ejemplos de castigos a no-castigadores son citados por Axelrod (1997):

Las metanormas se han usado en los sistemas de denuncia en las sociedades comunistas. Cuando las autoridades acusan a alguien de hacer algo malo, los demás son impulsados a denunciar al acusado. No unirse a esta forma de castigo es tomado en sí mismo como una deserción contra el grupo (Axelrod 1997, 72).

Axelrod (1986; 1997) cree que esta es una de las formas como se ejecuta una norma, a saber, no solamente castigando a aquellas personas que no la siguen, sino también castigando a aquellas personas que no hacen respetar, mediante el castigo, el seguimiento de dicha norma. Este investigador propone un modelo para estudiar la estabilidad de la cooperación, el castigo altruista y la *metanorma* bajo la perspectiva de la teoría de juegos evolutivos. El diagrama de la simulación llevada a cabo por Axelrod (1986) es el siguiente:



Figura 2.4. Diagrama de

*metanorma* en teoría de juegos evolutiva (Axelrod, 1986, 1101).

En este diagrama se tiene en cuenta no solamente la posibilidad o no de cooperar que posee un individuo, o la posibilidad que tiene de castigar la deserción, sino también la posibilidad que se presenta de castigar al que no ha castigado la deserción. Axelrod (1986) llevó a cabo una

simulación de cinco partidas cada una con veinte jugadores y una duración por partida de 100 generaciones. Sus resultados encontraron que al inicio de las interacciones el nivel de vengatividad (castigo) se incrementa, disminuyendo correlativamente el nivel de audacia (no-cooperación) hasta umbrales muy bajos. Sin embargo, los jugadores siguen manteniendo niveles altos de vengatividad, incluso aunque los *Free riders* sean escasos. Se puede decir que si un cooperador detecta una deserción y no la castiga, este comportamiento provocaría que otro cooperador no castigará una deserción también, y así sucesivamente hasta que el efecto cascada terminará por desplazar la cooperación, instaurando de nuevo un ambiente altamente egoísta dentro del grupo. La *metanorma* puede bloquear el efecto cascada al controlar las deserciones de segundo orden (no castigar a los desertores) con la amenaza de un castigo por no castigar. Esto se debe a que los jugadores permanecen atentos a castigar la deserción en cualquier momento del juego, pues siempre quieren evitar el castigo por no sancionar una deserción detectada. Para Axelrod (1986) este mecanismo provoca que el sistema sea auto-vigilante haciendo de la norma de reciprocidad y, por tanto, de la cooperación, una estrategia evolutivamente estable. En la estabilidad de dicha estrategia también están de acuerdo Boyd & Richerson (1992). Ahora bien, estos dos investigadores sostienen que hay dos formas cualitativamente distintas en las que la estrategia castigadora de segundo orden (el castigo a no cooperadores) puede reforzar la cooperación. En primer lugar, si los beneficios a largo plazo de la cooperación de un individuo que castiga son mayores que los costos en que incurre ese mismo individuo al aplicar la sanción, entonces las estrategias de: cooperar y castigar a los no cooperadores, de cooperar solamente si lo han castigado y si no siempre desertar y, en ocasiones, de cooperar pero no castigar coexisten en un equilibrio oscilatorio permanente, equilibrio polimórfico. Es decir, cada una de esas tres estrategias llega a turnarse el liderazgo en el número de individuos que la ejecutan dentro del grupo en largos periodos de tiempo. Pero, como ellos mismo sostienen el equilibrio se caracteriza por una mayoría de desertores y una minoría de las demás estrategias; la tasa de cooperación se mantiene en niveles bajos (Boyd & Richerson, 1992, 181). En segundo lugar, si el costo de ser castigador es suficientemente alto, la estrategia “moralista” o la *metanorma* como la denomina Axelrod (1986) puede ser evolutivamente estable dada la complementariedad de fuerzas entre dicha estrategia, el castigo altruista de primer orden y la Selección de grupos (Boyd & Richerson, 1992, 172-173).

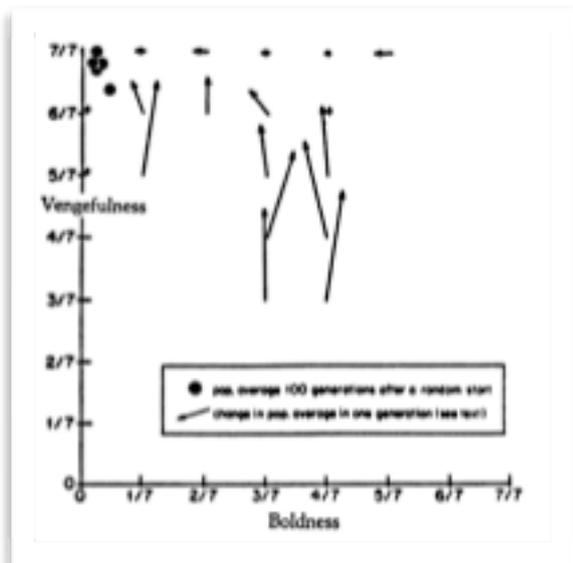


Figura 2.5. Cuando interviene una metanorma,

durante 100 generaciones, el incremento de la vengatividad es constante (dirección de las flechas) hasta el punto de hacer que la audacia se elimine y el sistema se mantenga auto-vigilado. (Axelrod 1986, 1102).

Como lo sostiene Boyd & Richerson (1992) y Axelrod (1986) lo importante de este mecanismo es que debe ser complementario con el de castigo altruista a los desertores, pues, si no se ligan ambos castigos la posibilidad de que un individuo deje de ser vengativo o castigador es alta, al ver que otros no han castigado una deserción (efecto cascada). Sin embargo, el modelo construido por Boyd & Richerson (1992) adolece, aunque con menor intensidad, del mismo problema presentado por los modelos de reciprocidad y castigo altruista de primer orden, a saber, el tamaño de los grupos incrementa el costo del castigo fortaleciendo los incentivos para los *Free riders* en todos los niveles de la cooperación. Según estos dos autores los modelos clásicos de castigo de primer orden tiene el problema de que al incrementar el tamaño del grupo también se incrementa, de forma exponencial, la expectativa en el número de interacciones necesarias para alcanzar la cooperación, pues el costo del castigo también es exponencialmente mayor, mientras que en su modelo, en donde se tiene en cuenta el castigo de segundo orden, este incremento es meramente lineal.

Con todo, la *Reciprocidad Fuerte*, que incluye el castigo de primer y segundo orden, requiere para su explicación de la existencia efectiva de una fuerza evolutiva a nivel de grupos que pueda explicar el hecho de por qué ciertas personas asumen costos individuales que benefician a los demás en interacciones sociales (Boyd & Richerson, 1992; Fehr, Fischbacher & Gächter, 2002;

Gintis et al, 2005). Pues, en cualquier caso los comportamientos castigadores estudiados hasta el momento no dejan de ser, desde el punto de vista biológico, comportamientos altruistas.

## ***2.7 Selección de grupos como explicación al altruismo***

Para Sober & Wilson (1998) es muy posible que el comportamiento altruista, señal de cooperación en los organismos biológicos, definido en la biología evolucionista como: el aumento de la aptitud de los demás y la disminución de la aptitud del actor (Sober & Wilson, 1998), haya podido presentarse gracias a la fuerza de la selección natural que se ejerce entre los grupos y no dentro de ellos. En otras palabras, según los autores, gracias al cumplimiento de ciertas condiciones naturales la fuerza de la selección al nivel de los grupos puede superar la fuerza al nivel individual, provocando la permanencia de un rasgo o conjunto de rasgos que reporten un costo para el individuo pero que, a su vez, presenten un beneficio para el grupo. En este sentido, Sober & Wilson (1998) defienden un pluralismo o mecanismo multinivel en los procesos de selección natural y bajo esta perspectiva un rasgo altruista puede prosperar aunque sea deletéreo para el individuo, pues puede ser beneficioso y adaptativo para el grupo.

Pareciera que un grupo que posee muchos individuos altruistas tiene baja aptitud frente a un grupo en donde predominan los individuos egoístas. Esto porque, según Sober & Wilson (1998), a menudo se considera que el grado de aptitud de un grupo se obtiene por el promedio de aptitud de los individuos que lo integran. Si esto fuera así, el promedio de aptitud de un grupo con muchos individuos altruistas, que por tener este rasgo, tiene baja aptitud individual, sería también bajo. Sin embargo, Sober & Wilson (1998) sostienen que un grupo con estas características puede tener una aptitud global mayor a pesar de tener individuos con una aptitud relativa menor en su interior. La razón de esta conclusión es que este es el típico caso de la *Paradoja de Simpson* o la *Falacia del Promedio* que indica que el promedio relativo de los individuos integrantes de un grupo no refleja acertadamente el promedio global del grupo, puesto que existe una contribución desigual a la media global del grupo; en tanto que los altruistas, aunque menos aptos, reportan a los egoístas y a los demás altruista beneficios y, por tanto, puntos adicionales de aptitud que no se ven reflejados en el promedio relativo, sino en la media global del grupo.

Para que el altruismo prospere en un grupo con egoístas que reciben beneficios de los altruistas y que, además, no incurren en ningún costo adicional por ayudar a otro, afirman Sober & Wilson (1998), existen ciertas condiciones que deben cumplirse, a saber:

1. “Debe haber más de un grupo, una población de grupos” (11), pues la selección de grupos se da cuando hay competencia entre ellos.
2. “Es necesario que los grupos sean *distintos* en cuanto a proporción de altruistas” (11), de tal suerte que cada grupo aporte o contribuya de forma distinta a la media global.
3. “Debe existir una relación directa entre la proporción de altruistas en el grupo y el rendimiento del grupo” (11).
4. Aunque por definición los grupos están aislados entre sí, en cierto sentido no están aislados (la progenie de ambos grupos debe mezclarse, o bien competir en la formación de nuevos grupos” (11). En el modelo de Sober & Wilson (1998) esta es la única posibilidad de que las ventajas relativas de los altruistas superen las ventajas relativas de los egoístas. Es decir, solamente al conjugarse la aptitud relativa de cada grupo en un solo grupo más grande se puede identificar el aporte relativo de cada uno y, sobre todo, la ventaja que los altruistas le aportan a la aptitud global del grupo. De otra manera al interior de cada grupo las ventajas relativas de los altruistas no podrán superar jamás las ventajas relativas de los egoístas.

En conclusión, Sober & Wilson (1998) argumentan que sí se dan estas condiciones en un medio natural y si, además, se tiene cuidado de no caer en la *Falacia del Promedio*, se podrá identificar cómo el diferencial de aptitud de los grupos es más fuerte que el diferencial de aptitud de los individuos para la permanencia de rasgos como el comportamiento altruista.

Es importante mencionar que el tipo de altruismo al que Sober & Wilson (1998) se están refiriendo en su modelo es el denominado por ellos como “altruismo extremo” (15), es decir, como lo expresan Fehr, Fischbacher & Gächter (2002), este tipo de altruismo puede denominarse como *incondicional*, pues el individuo que lo expresa no condiciona su comportamiento a la acción de los demás y, en este sentido, el altruismo dado no emerge como una respuesta al altruismo recibido. En otras palabras Fehr & Fischbacher (2005) sostienen que un altruista incondicional nunca ejercerá una acción que vaya en detrimento del pago de aquellos con quien interactúa. Sober & Wilson (1998) consideran, con todo, que este tipo de altruismo puede ser poco frecuente, raro en la naturaleza, dado que “a) la selección de grupos casi nunca es lo bastante fuerte para producir estos comportamientos o b) porque

habitualmente existe una manera de beneficiar al grupo sin este sacrificio extremo” (Sober & Wilson, 1998, 26).

Así pues, el modelo del altruismo incondicional no es una buena herramienta de explicación del comportamiento humano dado que es casi imposible no encontrar beneficios encubiertos en cualquier interacción, ya sea humana o no. Además, este modelo no puede explicar precisamente el carácter condicional de las relaciones sociales que evidentemente dominan nuestra vida diaria.

## ***2.8 La controvertida Selección de Grupos***

Desde su postulación inicial en 1962 por Vero C. Wynne-Edwards, la teoría de la Selección de grupos ha estado en constante debate. Algunos defensores como el filósofo australiano Kim Sterelny (2001; 2003) han extendido esta propuesta hacia la evolución de la cognición y la cooperación humana.

Gran parte del impulso a la Selección de grupos, y en general a la Selección multinivel, se ha ocasionado por la fundamental definición del llamado “Esqueleto lógico de la Selección natural” que propuso Richard Lewontin en su artículo de 1970, *The units of selection*. Allí, Lewontin sostiene que los requisitos para que exista evolución por selección natural deben ser tres:

1. Different individuals in a population have different morphologies, physiologies, and behaviors (phenotypic variation).
2. Different phenotypes have different rates of survival and reproduction in different environments (differential fitness).
3. There is a correlation between parents and offspring in the contribution of each to future generations (fitness is heritable). (Lewontin, 1).

En este sentido, Lewontin afirma que la variación heredable es eficacia biológica. Pero, además de esto, lo que entresaca este biólogo de su claridad lógica es que cualquier tipo de entidad que cumpla con estos tres requisitos puede ser denominada como “unidad de selección”. En otras palabras, los requisitos lógicos para que la selección natural se dé podrían teóricamente ser cumplidos por entidades de cualquier nivel, desde los genes hasta los superorganismos o grupos de organismos como lo proponen Sterelny & Griffiths (1999).

Debe anotarse, sin embargo, que la teoría de la Selección de grupos aún hoy en día tiene dificultades para ser aceptada en el mundo científico. Uno de los argumentos en contra de dicho mecanismo de selección es expuesto por el filósofo Michael Ruse en su libro *Mystery of mysteries. Is evolution a social construction?*. Él expone su argumento como sigue:

A group perspective has internal problems: although in the long run everyone might benefit, it is difficult to see why (other than in exceptional circumstances) a short-term devotion to self should not be preferred by selection. Adaptations directed toward immediate personal benefit would seem to be fitter than adaptations benefiting others, even if down the road all would gain with, and only with, the latter. Unfortunately, selection is necessarily a short-term process, without forethought for the future. (Ruse 1999, Pág, 129).

En este sentido, ¿cómo puede la selección natural emplear mecanismos que “piensen” o “consideren” ventajas y resultados adaptativos en etapas posteriores, si la materia prima de la que se vale la selección natural no es otra que las mutaciones aleatorias “de corto plazo” que se presentan en todo organismo? El rechazo a la Selección de grupos está encabezado por científicos tan prominentes como George Williams (1966; 1992), Lack (1966), Ghiselin (1974), Maynard Smith (1976), Dawkins (1982) o Gould (2002). Según Dawkins (1980), por ejemplo “un error común consiste en creer que la cooperación dentro de un grupo en un nivel dado de organización, debe aparecer mediante la selección de grupos [...] La teoría de la Estrategia Evolutivamente Estable ofrece una alternativa más austera” (1980, 360)

Además, muchos teóricos están de acuerdo en que la Selección de grupos, si llegase a existir, requiere de grupos pequeños que permanentemente se recombinen para permitir de esta forma que la fuerza de la Selección de grupos pueda realizar una contraposición efectiva a la fuerza de la Selección individual (Sober & Wilson, 1998; Fehr & Fischbacher, 2005; Gintis et al, 2005). Esta condición no ha sido lo suficientemente determinada en el marco de las relaciones sociales de los seres humanos. Además, como lo sostiene Nesse (1994, 633) la teoría de la Selección de grupos aún adolece de claridad conceptual y de falta de datos empíricos para su comprobación.

En cuanto a la claridad conceptual, filósofos y biólogos (Hull, 1980; Lloyd, 2001; Andrade & Fajardo, 2008) han intentado desarrollar aparatajes conceptuales bien definidos -y por que no, posturas ontológicas- que identifiquen con claridad elementos tan importantes para la selección

como son los de: la unidad de selección que difiere del nivel mismo de selección. Esto en palabras de Dawkins (1976) es, la identificación clara de el agente replicador y de la entidad que interactúa con el medio ambiente y a la que se le puede adscribir cambio adaptativo. A pesar de todos estos esfuerzos, la cuestión sigue en debate y controversia.

### 3. PROBLEMAS DE COORDINACIÓN EN LA TEORÍA POLÍTICA DEL ESTADO. CASTIGOS Y RECOMPENSAS

Como vimos en el capítulo anterior, la solución comúnmente aceptada en la actualidad como respuesta efectiva a los dilemas sociales de primer y segundo orden de la *Reciprocidad Fuerte*, entendida esta como el comportamiento relacionado con dos aspectos: i) la cooperación condicional y ii) el castigo altruista, tanto para los no-cooperadores como para los no-castigadores (Gintis et al, 2005; Fehr & Fischbacher, 2003 - 2005; Sethi & Somanathan, 2005; Bowles & Gintis, 2003; Fehr & Henrich, 2003; Fehr, Fischbacher & Gächter, 2002), se enfrenta a la dificultad de suponer, en este último aspecto, la necesidad de la Selección de grupos (aún hoy en día altamente controvertida dentro de la ciencia biológica), (Boyd & Richerson, 1992; Gintis et al, 2005; Gintis, 2008; Sethi & Somanathan, 1996; Fehr & Gächter, 2002; Fehr & Fischbacher, 2005).

El castigo altruista involucrado en la *Reciprocidad Fuerte* se ve explotado con distinta intensidad, dependiendo del tamaño del grupo y del nivel del castigo (castigo a no-cooperadores o a no-castigadores) por los individuos que prefieren no cooperar o castigar en este sentido, pero que de cualquier forma participan de los beneficios del sistema de castigos mantenido por otros (Boyd & Richerson 1992). Así pues, la forma de explicar la estabilidad de la *Reciprocidad Fuerte* y, por lo tanto, de solucionar los dilemas a los que se enfrenta, es sostener que la Selección de grupos es la razón por la cual un comportamiento no adaptativo a nivel individual, como es el castigo altruista, es viable en la naturaleza, en especial en el comportamiento humano (Boyd et al, 2005). Según esta interpretación, como ya lo vimos en el capítulo 2, la frecuencia del castigo y el comportamiento cooperativo es positivamente correlacionado a través de los grupos. De esta manera los grupos con una mayor cantidad de castigadores presentan ventajas competitivas frente a aquellos que tienen una tasa menor o que no tienen castigadores (Boyd et al, 2005, 217).

En lo que sigue abordaré la propuesta de Hobbes al problema del contrato social, que desde mi punto de vista, puede ofrecernos herramientas y respuestas más adecuadas para defender el papel del castigo en la consolidación efectiva de la cooperación, sin tener que asumir la hipótesis controvertida de la Selección de grupos. En este sentido, revisaremos la

interpretación, desde la teoría de juegos, que Jean Hampton (1986) elabora de las tesis hobbesianas para lograr una mejor comprensión del papel del castigo en la cooperación.

### ***3.1 Egoísmo, guerra y racionalidad en Hobbes***

El problema que Hobbes aborda desde el inicio de su *Leviatán* (1651), es el de buscar los fundamentos de la ley, el derecho y el gobierno en la propia naturaleza humana para alejarse de la postura imperante de la época, que consideraba al derecho divino como origen de la regulación social y era fundamentado en principios externos al individuo y alejados de su propio interés. Desde este punto de vista, lo que pretende Hobbes, puede decirse, es elaborar un experimento mental en donde se identifique, por un lado, la profunda naturaleza humana que nos enfrenta a permanentes dificultades cuando vivimos en agrupación y, por el otro, determinar los mecanismos y procesos que permiten a los individuos, en contra de las dificultades naturales, conformar comunidades estables en donde se privilegia el derecho a la propiedad privada y el bienestar comunal por encima de los conflictos iniciales generados en un hipotético estado de naturaleza.

Así pues, para Hobbes el estado de naturaleza es anterior al estado civil o político y se caracteriza porque en él no existe ninguna instancia superior de control, tanto normativo como punitivo. Hobbes lo describe así: “cuando no existe un poder capaz de imponerse a todos ellos [los hombres en libertad]” (XIII. 102/95<sup>7</sup>). De la misma forma, sostiene que en esta condición “no existe propiedad ni dominio, ni distinción entre *tuyo* y *mío*; solo pertenece a cada uno lo que puede tomar, y solo en tanto que puede conservarlo” (XIII. 104/98). Esta situación hipotética que Hobbes configura, le sirve para justificar el poder político al que los hombres deben llegar por mutuo acuerdo para superar las errores del estado de naturaleza.

En concreto, las fallas que presenta el estado de naturaleza en el mantenimiento de la propiedad y el aseguramiento de la convivencia pacífica de los individuos obligan a que estos prefieran superar dicho estado mediante la constitución de un poder capaz de imponerse a todos. Así, un estado político o civil, como lo denomina Hobbes, surge de los conflictos del estado natural, pues la principal función de aquel es eliminar los defectos de este último. Con todo, el paso de

---

<sup>7</sup> Señalaré la paginación del *Leviatán* en este orden: primero la paginación de la versión en español (Hobbes, T. *Leviatán*. México: Fondo de Cultura Económica. 1980). Segundo la paginación de la versión en inglés (Hobbes, T. *Hobbes's Leviathan*. Oxford: Oxford University Press. 1909).

un estado a otro no es un proceso que Hobbes considere obvio y natural. Por el contrario, para “subir los escalones” es necesario que los hombres lleguen a uno o varios acuerdos, de tal forma que puedan liberarse de una profunda situación de conflicto y guerra en donde su autopreservación se ve en constante amenaza. Recordemos que para Hobbes el interés primordial del ser humano es el interés egoísta, pero muy natural, de la autopreservación, incluso por encima de la preservación de otros (Hobbes XIII. 101/95).

Esta transición de un estado a otro es, efectivamente, el desarrollo de las tesis liberales que sostienen que la justificación de toda autoridad política y de gobierno debe radicar en el consentimiento de los individuos. Es el principio de legitimidad política que, desde los preceptos liberales, ya Hobbes expresa en su *Leviatán*. Pero antes de abordar el paso específico de una condición de guerra permanente a una de convivencia bajo un poder absoluto, es bueno aclarar las razones que hacen que Hobbes sostenga que un estado en donde no hay autoridad suprema es, por definición, un estado en guerra.

### **3.1.1 Inferencia hobbesiana del estado de guerra**

Hobbes afirma que en el estado de naturaleza todos los hombres son aproximadamente iguales en cualidades, disponen de las mismas capacidades físicas e intelectuales (XIII. 100/94). Las diferencias que pueden darse entre ellos no son significativamente relevantes como para que alguno de ellos pueda destacarse sobre los otros y dominar a su antojo (XIII. 100/94). Dada esta igualdad aproximada, adicionalmente todos los hombres poseen las mismas expectativas relacionadas con alcanzar sus fines e intereses. El problema tiene que ver con que los fines de los individuos son eminentemente egoístas, por tanto, no tienen en cuenta fines comunales o altruistas. En el estado de naturaleza cada quien tiene que arreglárselas para lograr satisfacer sus intereses individuales sin importar los intereses ajenos o comunales. En este ambiente todos los hombres tienen derecho a todas las cosas, aunque ningún derecho exclusivo sobre ninguna de ellas. Bajo esta perspectiva, los hombres se verán inevitablemente enfrentados al momento de realizar sus fines, tanto así que en ocasiones, para Hobbes muy comunes, sus propias vidas se verán amenazadas por aquellos que los consideran obstáculos para satisfacer sus intereses propios. Así pues, gracias al egoísmo enraizado en la naturaleza humana y sus consecuencias en el desarrollo de los intereses particulares en competencia con otros, podemos entender la famosa sentencia de Hobbes: *homo homini lupus*.

El inevitable conflicto entre individuos en libertad que pueden emplear estrategias y artimañas de toda clase hace que en el hombre, según Hobbes, se despierte un profundo temor hacia los demás. Esto se debe a que, dado que todos los hombres tienen la potestad de alcanzar sus fines por los medios que ellos crean convenientes, probablemente cuando dos o más deseen lo mismo la enemistad surgirá entre ellos y al convertirse en competidores y obstáculo para la satisfacción del interés egoísta, tratarán de hacerse daño y hasta eliminarse mutuamente. Esto despierta en cada hombre un consecuente temor frente a los demás y una profunda desconfianza hacia las intenciones ajenas.

De aquí que un agresor no teme otra cosa que el poder singular de otro hombre; si alguien planta, siembra, construye o posee un lugar conveniente, cabe probablemente esperar que vengan otros, con sus fuerzas unidas, para desposeerle o privarle, no solo del fruto de su trabajo, sino también de su vida o de su libertad. Y el invasor, a su vez, se encuentra en el mismo peligro con respecto a otros. (Hobbes XIII. 101/95).

Dada esta profunda desconfianza, es natural y hasta necesario que los hombres sojuzguen a los demás para evitar ser dañados. Esta actitud de violencia preventiva -anticipación, en términos de Hobbes- tiene que ser mantenida por los individuos hasta que no exista ningún poder capaz de amenazar su vida y su propiedad (Hobbes XIII. 101/95). Como se deriva de los presupuestos hobbesianos, dicha actitud está racionalmente justificada por la preocupación permanente de los individuos de su autopreservación, en medio de un estado de cosas en donde “todo vale” para cumplir con los fines particulares. De aquí, pues, se colige que en un estado de naturaleza el hombre se encuentra en un permanente estado de guerra con los demás, dada su principal finalidad de autoconservación y sus profundos sentimientos de temor y desconfianza frente a los demás. Hobbes sostiene que una situación de guerra tal es una mera *inferencia* que podemos hacer teniendo en cuenta las pasiones del hombre (XIII. 103/97). Recordemos que este estado de naturaleza y el homólogo estado de guerra son una construcción meramente teórica de la reflexión hobbesiana.

Ahora bien, el estado de guerra para Hobbes no es en modo alguno una situación bélica total en donde los individuos están exclusivamente dedicados a la lucha. Es más bien un ambiente de tensión constante frente a los demás, fundamentado en la desconfianza y el temor. Escuchemos a Hobbes: “la naturaleza de la guerra consiste no ya en la lucha actual, sino en la disposición

manifiesta a ella durante todo el tiempo en que no hay seguridad de lo contrario. Todo el tiempo restante es paz”. (XIII. 102/96).

Kavka (1986) sostiene que la concepción hobbesiana del estado de guerra se entiende mejor al contrastarla con la concepción delineada por Locke en su *Segundo ensayo del gobierno civil*. Según Kavka para Locke el estado de guerra se da cuando existe una declaración explícita de un individuo hacia otro que transmite su intención de enemistad y destrucción. Por el contrario, para Hobbes no es necesario tal signo explícito de hostilidad, pues en el estado de naturaleza siempre hay una disposición de enemistad constantes, a menos que se presenten al individuo garantías de lo contrario.

La lamentable consecuencia de una situación como la descrita por Hobbes en el estado de naturaleza tiene que ver con que bajo dicho estado de cosas no se puede pensar en ningún desarrollo personal y, mucho menos, colectivo que permita alcanzar mayores beneficios en todos los aspectos. Hobbes enumera algunos desarrollos deseables, pero impracticables en un estado de constante guerra:

En una situación semejante no existe oportunidad para la industria, ya que su fruto es incierto; por consiguiente no hay cultivo de la tierra, ni navegación, ni uso de los artículos que puedan ser importados por mar, ni construcciones confortables, ni instrumentos para mover y remover las cosas que requieran mucha fuerza, ni conocimiento de la faz de la tierra, ni cómputo de tiempo, ni artes, ni letras, ni sociedad; y lo que es peor de todo, existe continuo temor y peligro de muerte violenta; y la vida del hombre es solitaria, pobre, tosca, embrutecida y breve. (Hobbes XIII. 103/97).

Es importante mencionar que para Hobbes, el papel desempeñado por la razón en medio de este frenesí de pasiones, intereses y conflictos es uno meramente instrumental. Podría decirse, como lo sostiene Taylor (1994, 40) que la razón en Hobbes es utilitaria, pues se emplea en pro del logro de los fines e intereses de los individuos, surgidos de sus propias pasiones.

Resumiendo, la inferencia del estado de guerra elaborada por Hobbes puede expresarse en los siguientes términos:

1. Los hombres son por naturaleza aproximadamente iguales en cualidades. (Hobbes XIII. 100/94).
2. La pasión dominante en los seres humanos es el deseo de autopreservación. (Hobbes XIII. 101/95).
3. Todos los hombres tienen la misma esperanza de satisfacer sus deseos y alcanzar sus fines. (Hobbes XIII. 101/95).
4. Por tanto, se genera entre los hombres competencia por lograr sus fines e intereses (el principal de los cuales es la autopreservación a toda costa). (Hobbes XIII. 101/95).
5. La competencia es causa de lucha, enemistad y guerra. (Hobbes XI. 80/75).
6. El temor y la desconfianza generados por las intenciones egoístas de los hombres los invade, provocando que la mejor forma de lograr la autopreservación sea sojuzgar anticipadamente a los demás en la medida en que ellos son potenciales enemigos. (Hobbes XIII. 101/95).
7. En conclusión, en el estado de naturaleza la seguridad es un bien escaso. El estado de naturaleza es un estado de constante guerra. (Hobbes XIII. 104/98).

Adicionalmente, Hobbes asegura que existe otro factor que genera conflictos en el estado de naturaleza, pero que en estricto sentido no es fundamental para la inferencia. Este factor es la gloria que sienten los individuos y que los obliga a exigir de los demás el reconocimiento, hacia sus logros, ventajas y capacidades. Para Hobbes esta pasión refuerza el conflicto, pues, por un lado, los individuos siempre estarán buscando satisfacer sus deseos y sobresalir en la comparación con otros, de tal forma que la competencia es cada vez más necesaria. Por otro lado, el deseo de gloria hace que las relaciones entre individuos sean tensas, pues cada cual espera y exige que el otro alabe y reconozca todo logro y capacidades propias. Para Hobbes, aquellos que sin tener dichas capacidades esperan de los otros un tal respeto y reconocimiento sobre algo inexistente, se encuentran guiados por una pasión más vergonzosa, la vanagloria.

Así hallamos en la naturaleza del hombre tres causas principales de discordia. Primera, la competencia; segunda, la desconfianza; tercera, la gloria. [...] la tercera, recurre a la fuerza por motivos insignificantes, como una palabra, una sonrisa, una opinión distinta, como cualquier otro signo de subestimación, ya sea directamente en sus personas o de modo indirecto en su descendencia, en sus amigos, en su nación, en su profesión o en su apellido. (Hobbes XIII. 102/96).

### 3.1.2 El dilema del estado de naturaleza

Como hemos visto, Hobbes configura un hipotético estado de naturaleza poblado por las pasiones de los hombres en igualdad de fuerzas, quienes quieren a toda costa mantenerse vivos, satisfaciendo sus intereses, objetivos y apetitos; incluso, en contra de la preservación de los demás. Todo esto se desarrolla mediante la aplicación de una racionalidad utilitaria e instrumental que los empuja a tomar decisiones que generan más conflicto (vg., las medidas de anticipación para sojuzgar a los otros). Es fácil ver en esa situación descrita por Hobbes un clásico Dilema del Prisionero (DP), en donde la racionalidad egoísta lleva a los hombres a tomar decisiones evidentemente negativas frente a sus propios intereses y a su bienestar. El capítulo XIII del *Leviatán* propone que el conflicto prevalece por sobre la cooperación gracias a la intensa búsqueda de la satisfacción de los intereses particulares de cada persona. Búsqueda que los empuja hacia la desconfianza y el temor frente a los demás y les muestra como única salida la violencia preventiva para salvar sus propias vidas.

Así pues, la característica para resaltar y poder interpretar el estado de naturaleza como un DP es la relacionada con la desconfianza y el temor mutuos que provocan inseguridad en dicho estado y que, desde un punto de vista racional, obliga a que la estrategia obvia para los hombres sea la de la anticipación (violencia preventiva). Ahora bien, el DP se configura con la estrategia de la anticipación, racionalmente adoptada, que es precisamente la estrategia que en definitiva provoca el estado de guerra (paso seis (6) de la inferencia hobbesiana). La razón lleva a los individuos a escoger una estrategia poco benéfica, tanto individual como colectivamente. Veamos porque es racional la estrategia de la anticipación y, por ende, bajo una lógica del clásico DP, inevitable un estado de guerra.

Dos individuos en el hipotético estado de naturaleza de Hobbes, Atticus y Baltasar, han logrado en el transcurso de sus vidas algunos bienes y, también, satisfacer sin problemas sus intereses particulares. Sin embargo, cada uno de ellos desea bienes que el otro posee. En tanto nadie tiene derecho exclusivo a las cosas, Atticus y Baltasar albergan la esperanza de poseer tales bienes en algún momento. Así pues, ellos pueden seguir dos estrategias, a saber: a) mantener un comportamiento pasivo frente al otro y, si es necesario, defenderse de los ataques del otro, o b) asumir una actitud de ataque, *anticipando* el probable comportamiento belicoso del otro, e invadir o tomar el bien deseado a través de la violencia. Es evidente que si ambos individuos permanecen pasivos, sus bienes no estarán en peligro, los mantendrán bajo su poder, a menos,

claro, que un tercero los pretenda también, con lo cual se enfrentarán nuevamente a un clásico DP. Sin embargo, dada la gran ventaja que puede reportar para ellos la sorpresa (anticipación) si uno de ellos ataca, mientras el otro permanece en una actitud pasiva, es muy probable que obtenga lo deseado sin mucha dificultad ni riesgo. Además de la ventaja que esta estrategia le ofrece a los individuos, es importante tener en cuenta que también es racional, como lo sugiere Hobbes (XIII. 101/95), anticipar o atacar por sorpresa, si se teme y desconfía de las intenciones del otro. En esta última situación, el atacado por sorpresa (es decir, el que mantenía una actitud pasiva) no solamente perderá la posesión en conflicto, sino que además está claramente expuesto a perder su libertad o hasta su vida en manos del atacante sorpresivo o anticipador. Por otro lado, si en cambio ambos se atacan, siguiendo la estrategia de la anticipación, los bienes y la vida de los dos corren peligro. Pero incluso así esto es mejor que terminar esclavizado o asesinado por sorpresa. Por lo menos, en este caso la lucha predeterminada y planeada abre la posibilidad de hacer daño al otro y de recibir menos daño; más de lo que se puede esperar defendiéndose frente a un ataque sorpresivo. Veamos la matriz de ganancias que una situación como la del estado de naturaleza nos ofrece:

Tabla 3.1. Matriz de ganancias Estado de Naturaleza.

		B	
		Pasivo	Anticipado
A	Pasivo	3;3	1;4
	Anticipado	4;1	2;2

Como se ve, esta matriz es idéntica a la matriz que arroja un típico y clásico DP, en donde la estrategia dominante es desertar, para nuestro caso, anticipar. Pues sin importar la decisión del otro, cualquiera de los dos individuos racionalmente optarán por no cooperar (anticipar, invadir o atacar), pues es la estrategia dominante. Más específicamente, si Baltasar mantiene una actitud pasiva (cooperar), será mejor para Atticus atacar, pues con ello obtiene una ganancia mayor que permanecer idénticamente pasivo (4;1). De la misma forma, si Baltasar ataca o, en términos de Hobbes, anticipa, sigue siendo mejor para Atticus anticipar, pues el daño será menor que no hacerlo (2;2). Recordemos que este tipo de juegos se caracterizan por su simetría,

es decir que ambos jugadores disponen de las mismas opciones y cada una de ellas ofrece los mismos resultados para cada cual.

Si los individuos reflexionan sobre sus posibilidades y maximizan su propia preservación, racionalmente ambos tendrán que decidir no cooperar, esta es la opción (2;2) que se constituye en un Equilibrio de Nash como ya lo expliqué en el primer capítulo (Sección 1.1.). Ahora bien, como se nota en la matriz, este Equilibrio de Nash no es para nada la mejor situación que ambos individuos podrían lograr, pues, la de mantenerse igualmente pasivos expresa una ganancia de (3;3), es decir, la mejor para ambos a la vez. Esta es la situación Pareto eficiente que no se alcanza con la estrategia dominante de la anticipación. Con todo, en el estado de naturaleza la situación más aceptable desde un punto de vista colectivo, la de cooperar al unísono, es la más difícil de alcanzar, dados los temores y las desconfianzas frente al otro que obligan a que la estrategia racional, desde un punto de vista individual, sea la de no cooperar. Esta es, como lo afirma Hampton (1986), la explicación racional del conflicto, puesto que el estado de guerra es explicado mediante las dinámicas racionales de los individuos en un estado de naturaleza.

### **3.1.3 Acuerdo para salir del estado de naturaleza**

Como vemos, Hobbes justifica la irracionalidad de la cooperación en un estado de naturaleza. Y , por supuesto, justifica también la racionalidad del no respeto de los pactos, si se dan en este hipotético estado de cosas. En este sentido, Hobbes sostiene que:

Quando se hace un pacto en el que las partes no llegan a su cumplimiento en el momento presente, sino que confían una en otra, en la condición de mera naturaleza (que es una situación de guerra de todos contra todos) cualquier sospecha razonable es motivo de nulidad. [...] En efecto, quien cumple primero no tiene seguridad de que el otro cumplirá después. (Hobbes XIV. 112/105).

Se entiende, por la aseveración de Hobbes, que la misma racionalidad que obliga a atacar o invadir a los otros (racionalidad expresada en el oportunismo y provocada por el miedo y la desconfianza), es también la que obliga a no cumplir con los pactos y convenios en medio de un estado de naturaleza como el descrito. Esto, porque la desconfianza, que aparece aquí como emoción protagónica, obliga a que, mediante la reflexión, los individuos no esperen de los otros el cumplimiento de lo acordado. Por tanto, no es racional ser el primero en cumplir. Al

incumplir se evita ser el tonto o ser explotado con facilidad. En el caso de los pactos nuevamente aparece un DP clásico, es decir uno de una sola jugada.

Ahora bien, junto con Hampton (1986), podemos preguntar ¿cómo es posible que hombres libres en el estado de naturaleza puedan ponerse de acuerdo y convenir la institucionalización de un soberano absoluto, con el suficiente poder para limitar sus libertades y obligarlos a cooperar, si para ellos es racional no cooperar y, además, es racional no respetar los pactos, acuerdos y convenios?

Una posible respuesta a esta objeción para instaurar un soberano absoluto a partir de un estado de naturaleza, es la que han esgrimido algunos estudiosos, a saber: que es necesario considerar la interacción de los hombres libres inmersos en un estado de naturaleza como participantes de un Dilema del Prisionero Iterado (DPI). Esta, evidentemente es una consideración más realista y, por supuesto, una salida a la dificultad de no poder hacer pactos en una condición de mera naturaleza. Como ya lo vimos en el capítulo anterior (Sección 2.3.), la dinámica de un DPI permite considerar a la cooperación como una estrategia racional, puesto que los individuos preferirán cooperar con su contraparte dada la posibilidad, muy segura, de futuras e indefinidamente repetidas interacciones con esa misma persona. En tal caso, los individuos prefieren cooperar para mantener buenas relaciones (buena reputación) que les permita seguir cooperando con las personas con las que repetidamente se van encontrando en su estado de naturaleza. Es evidente, que en este caso de DPI, la cooperación está impulsada no solamente por los cálculos racionales del individuo que ve futuras cooperaciones y beneficios con las mismas contrapartes, sino que también -y por supuesto, empleando la racionalidad instrumental- por las posibles retaliaciones o castigos que pueda recibir si en repetidas interacciones sus contrapartes han sido defraudados por su actitud egoísta. Veamos la matriz de preferencias de un DPI, que viene a ser el escenario más realista en el que estarían inmersos los individuos hobbesianos:

Tabla 3.2. Matriz de ganancias del DPI

		B	
		Cumplir	Incumplir
A	Cumplir	4;4	1;3
	Incumplir	3;1	2;2

Hampton (1986) sostiene que el DPI tiene la misma matriz que el juego propuesto por Amartya Sen y que denominó en su momento Juego de Confianza (*Assurance Game - AG*). Aunque es cierto que la matriz es la misma, no estoy de acuerdo con Hampton en equiparar ambos juegos. Como más adelante veremos, los problemas de cooperación como el DP o el DPI son analíticamente distintos a los problemas de coordinación como el AG (*cf.*, Sección 3.2.).

Ahora bien, analizando la matriz de DPI, encontramos que si algún individuo incumple el pacto, a pesar de llevarse los beneficios del trato sin tener que entregar nada a cambio, su acción dañará su reputación frente a la contraparte y frente a los potenciales socios de futuros tratos (3;1). Así pues, como se ve, incumplir mientras el otro cumple no es la mejor opción, pues, en todo caso la propia imagen ante los demás se ve afectada. La mejor opción será, por supuesto, que ambos cooperen, pues en este caso la ganancia de la ayuda mutua se dará y, además, se fortalecerá la reputación para futuros tratos (4;4). Es importante resaltar que la característica relevante de este tipo de juegos es que no hay, como en el DP, una estrategia dominante, una que los jugadores adopten sin importar qué hace el otro. Si un jugador coopera, lo mejor que puede hacer el otro es cooperar, pero si un jugador incumple y no coopera, lo mejor que puede hacer el otro es no cooperar. Para el caso del DPI la explicación tiene que ver con que el beneficio mayor se da cuando ambos cooperan, y coincide con el beneficio de largo plazo (si se suma el beneficio bruto más la reputación construida que facilitará otros tratos exitosos).

Con todo, aparece ante nosotros una nueva cuestión. Como vemos, considerando un estado de naturaleza en donde las personas se encuentran inmersas en un DPI, es comprensible que sea racional cooperar con los demás y, si esto es así, entonces ¿por qué es necesario un contrato social?, ¿por qué es necesario un soberano absoluto que obligue a las personas a cooperar por su propio bien?

Es el momento de recordar lo dicho por Hobbes:

En efecto, quien cumple primero no tiene seguridad de que el otro cumplirá después, ya que los lazos de las palabras son demasiado débiles para enfrentar la ambición humana, la avaricia, la cólera y otras pasiones de los hombres, si estos no sienten el temor de un poder coercitivo; poder que no cabe suponer existente en la condición de mera naturaleza. (Hobbes XIV. 112/105).

¿Qué situación debemos considerar para que, a pesar de una dinámica de DPI, incluso así, los hombres creen que es más racional no cooperar y de esta manera encuentren necesario instaurar un soberano absoluto? Preliminarmente y a simple vista, parecería ser una situación llena de contradicciones. Pues, por un lado, es una situación en donde la cooperación puede darse a partir de la reflexión de las posibilidades de interacciones futuras, pero, por otro lado, existe algo que parece hacerles creer a las personas que es irracional cooperar. Por si esto fuera poco, en esta situación hay que considerar que a pesar de la creencia de estar sumergidos en un DPI, los individuos en un tal estado pueden llegar a un máximo acuerdo en donde se instaure un soberano absoluto, que tenga como tarea la de limitar las libertades individuales para fomentar los beneficios colectivos.

Aquí Hampton elabora una propuesta muy atractiva que puede dar cuenta de estas dificultades. En su libro de 1986 *Hobbes and the Social Contract Tradition*, Hampton explora las dificultades que desde la teoría hobbesiana del estado, expuesta en el *Leviatán*, pueden generarse para la consolidación de un Contrato social, e intenta algunas vías de solución a partir de ciertos conceptos desarrollados por la teoría de juegos.

Hampton sostiene que dada la psicología característica de los individuos hobbesianos, a saber, que los seres humanos son eminentemente egoístas, pues sus premisas de acción tienen que ver con su supervivencia a costa de cualquier cosa, incluso a costa de la supervivencia de los demás, es francamente difícil que la persona que sobrevivan en un estado de naturaleza -estado de virtual conflicto y guerra con los demás- puedan alcanzar un acuerdo para vivir en comunidad bajo las órdenes y el poder de un soberano; cediendo, por tanto, su derecho individual a todas las cosas.

Hampton (1986) configura su propuesta de interpretación de este problema, argumentando que los individuos hobbesianos, dado su carácter egoísta, han de sufrir de *miopía cognitiva* (no pueden ver los beneficios a largo plazo que la cooperación les promete). Esta sería la situación que buscamos para explicar un estado de cosas, en apariencia, contradictorio. Así pues, en tanto el auto-interés es el principal y mayor objetivo, muy probablemente los individuos nunca o en muy pocas ocasiones colaborarán entre sí. Pues, al concentrarse en las ventajas y beneficios de corto plazo que puede dar la estrategia de explotar y aprovecharse de los demás, no verán las ventajas y posibilidades de largo plazo que la cooperación les ofrece.

Para Hampton los individuos en el estado de naturaleza pueden exhibir dos tipos de *miopía cognitiva* (88). La primera tiene que ver con la incapacidad que algunos individuos pueden exhibir al desestimar los beneficios de largo plazo por una sobreestimación de los beneficios a corto plazo. Hampton sostiene que aquellos individuos que siempre o casi siempre no alcanzan a apreciar los beneficios de largo plazo que una situación les puede asignar, son evidentemente idiotas o cortos de entendimiento (82). En este primer caso, que podríamos denominar *miopía cognitiva de primer orden*, el individuo no cooperará con nadie porque no percibe o desestima el beneficio de esta acción para sus propios intereses. El segundo tipo de miopía, *miopía cognitiva de segundo orden*, se relaciona con una sobrestimación de la existencia de la miopía ajena. Es decir, el individuo se distancia de la cooperación porque cree que es altamente probable encontrarse con una persona miope en la interacción y, por tanto, es altamente probable que en lugar de recibir beneficios por su cooperación, esto vaya en contravía con sus intereses.

La primera clase de *miopía* parece estar apoyada por la afirmación, ya vista, de Hobbes, en cuanto a que las palabras son muy débiles para enfrentar la avaricia y la ambición humana (Hobbes XIV. 112/105). Es decir, que las promesas de un bien futuro, o el bien futuro mismo, nos son tan fuertes como para resistir las pasiones generadas por la posibilidad de recompensas inmediatas o de corto plazo. La segunda clase, también puede verse en las afirmaciones de Hobbes, por ejemplo cuando sostiene que: “Cuando se hace un pacto en el que las partes no llegan a su cumplimiento en el momento presente, sino que confían una en otra, en la condición de mera naturaleza [...] cualquier sospecha razonable es motivo de nulidad” (Hobbes XIV. 112/105). En este sentido, las personas no confían en las buenas intenciones de los demás, creen que los demás no están interesados en la cooperación, pues, como diría Hampton, unos están ciegos ante los beneficios futuros, por causa de las ganancias inmediatas y otros sobreestiman esa ceguera en los demás, tanto que abundan las sospechas para anular cualquier contrato. Así pues, según Hampton, la mayoría de los pactos en una condición de mera naturaleza son poco confiables por causa de la *miopía cognitiva* de las personas. Por tanto, mientras sea tan alta la probabilidad de que la contraparte actúe irracionalmente, miope o cortoplacista, será, a su vez, irracional la elección de una estrategia cooperativa en la consolidación de convenios y tratos (Hampton, 73).

Lo que sostiene Hampton es que la *miopía cognitiva de primer orden* es un defecto en el razonamiento de las personas que no ven (mediante la reflexión) que los beneficios futuros son

mayores que los de corto plazo (Hampton 82). Asumen equivocadamente, es decir: a partir de creencias erróneas, que la posibilidad de futuros encuentros con sus contrapartes es “remota”, muy poco frecuentes (Hampton 82). Esto les hace inferir que se encuentran en un DP clásico de una sola jugada y, por lo tanto, comprenden que la estrategia dominante es la de desertar en la interacción. Para Hampton este defecto de razonamiento es común en el estado de naturaleza hobbesiano, pero no por eso, ubicuo. Por su puesto, aquellos individuos a los que no les afecta, o no a menudo, tal defecto de razonamiento, siempre se verán cohibidos para cooperar (temerosos y desconfiados), pues consideran que los demás sí actúan bajo dicha visión cortoplacista. Esta última actitud, afirma Hampton, es nuevamente la muestra de la *miopía cognitiva*, pues ahora, con creencias falsas, se puede estar exagerando la miopía de los otros.

La gloria, pasión que esgrime Hobbes en el capítulo XIII como una de las causas del conflicto, es, siguiendo a Hampton (86), un factor que refuerza el conflicto y puede considerarse como una de las causas de la *miopía*; es decir, causa de ese defecto de razonamiento. La gloria, entendida como una posición triunfalista frente a la realidad hace que los individuos se sobreestimen y consecuentemente se crean más fuertes y capaces de lo que realmente son individualmente, esto genera en ellos una sobrevaloración de los resultados a corto plazo, frente a benéficos futuros logrados en comunidad. La vanagloria (posición triunfalista sin motivos), además, puede generar actos torpemente egoístas, causantes de más desconfianza (Hampton 87). “[...] el hombre, cuyo goce consiste en compararse a sí mismo con los demás hombres, no puede disfrutar otra cosa sino lo que es eminente.” (Hobbes XVII. 139/130).

Hampton afirma - siguiendo a Hobbes - que la situación de *miopía* es la razón por la que el conflicto puede generarse en un estado de naturaleza. O, en otras palabras, constituye los elementos que les hacen creer a los individuos del estado de naturaleza que se encuentran inmersos en un DP clásico y no en un DPI. Así pues, podemos entender por qué para Hobbes la solución a dicho conflicto ha de ser la instauración de un soberano absoluto que induzca a la cooperación generalizada mediante la eliminación de la *miopía cognitiva*. Dicha eliminación es efectuada por el soberano a través de la aplicación de incentivos y sanciones. En otras palabras, el papel del soberano absoluto será hacer que los individuos cooperen bajo sus órdenes, presentándoles incentivo que les permitan adquirir beneficios inmediatos o a mediano plazo de una vida en comunidad (*p.e.*, protección, generación de mercados, construcción de infraestructura, etc.). Además, su objetivo también es infringir castigos inmediatos a la gente

que no coopere; castigos, que en el estado de naturaleza los individuos veían muy lejos en el tiempo y por tal motivo los desestimaban.

Si esto es así, y la propuesta de Hampton es adecuada para explicar el estado de naturaleza de Hobbes, nuevamente: ¿cómo es posible que los individuos puedan llegar a un gran acuerdo para instaurar a un soberano absoluto? Hampton es consciente de esta dificultad (Hampton 73). Lo que hará es precisamente solucionar este problema en la teoría hobbesiana, identificando el tipo de acciones que deben llevar a cabo los individuos en un estado de naturaleza para cooperar, por lo menos por una única vez, y así instaurar al soberano y acceder a un contrato social.

Ahora bien, según Hampton (85), dado que el estado de naturaleza no es un estado de permanente conflicto - como lo aclara Hobbes (XIII. 102/96) - puede aparecer la cooperación entre individuos en situaciones de bajo riesgo. Piénsese que la *miopía de primer orden* no es ubicua en la población de un estado de naturaleza, por tanto, en ocasiones pueden darse convenios sin la interferencia de este error de razonamiento. Es fácil suponer que estos pocos convenios se darán principalmente en situaciones de bajo riesgo para los individuos involucrados en el pacto. Es decir, que en una situación como la del DPI puede ser racional cooperar en escenarios que reporten muy poco riesgo para los agentes que interactúan. Sin embargo, teniendo en cuenta la *miopía*, en situaciones de alto riesgo como puede ser la instauración de un soberano absoluto y, por tanto, la cesión de gran parte de sus derechos a todas las cosas, los individuos siempre actuarán bajo el marco del DP de una sola jugada en donde es racional no cooperar.

En este sentido, Hampton intenta desvirtuar la interpretación de esta situación como un DP de una sola jugada, proponiendo un marco diferente de acción que permita la instauración del soberano y la generación de un gobierno. Dicha solución radica en el carácter mismo de la psicología hobbesiana, pues los individuos no operan a partir de incentivos éticos (de ser así, podrían fácilmente acordar la cooperación sin requerir de un soberano), sino que por el contrario, lo hacen por incentivos de auto-interés (Hampton, 135). Hampton propone que la situación del paso de un estado de naturaleza a un acuerdo de instauración de soberano o contrato social se debe enmarcar en un acuerdo de auto-interés por parte de los individuos hobbesianos. Estos acuerdos se cuentan dentro de los problemas de coordinación en la teoría de juegos. La característica principal de los problemas de coordinación es que en las situaciones

enmarcadas por ellos las decisiones de los agentes son interdependientes, presentándose, de esta forma, paridades o coincidencias en los intereses de los individuos (lo que en teoría de juegos se denomina tener dos puntos de equilibrio o dos equilibrios de Nash). Estos puntos de equilibrio indican que las acciones de los individuos al coordinarse pueden dar mejores resultados que al no hacerlo.

Es importante mencionar aquí que Hampton acepta dos formas generales de solución a los problemas de coordinación. En primer lugar – sostiene –, pueden superarse los problemas de coordinación mediante la presencia de incentivo (negativo o positivos) en la situación problemática que permitan que las preferencias de las partes sean guiadas a uno de los puntos de equilibrio presentes en dicha situación (138). En este sentido, los incentivos (premios o castigos) pueden hacer que los individuos se hagan conscientes de la relevancia de ciertos puntos de equilibrio que benefician la cooperación. Así pues, los premios o castigos pueden hacer que las partes identifiquen la cercanía entre sus preferencias individuales para lograr una empresa comunal. En segundo lugar, Hampton sostiene que se pueden superar los dilemas de la coordinación de preferencias a través de la construcción de ambientes altamente informados que generarían posibilidades de acuerdos de auto-interés (138). En este caso, mediante la información y comunicación compartidas los individuos pueden identificar con alguna certeza las preferencias de la contrapartes, de tal forma que con esta información puedan coordinar sus propias preferencias para el logro de un bien común.

Así pues, Hampton argumenta que dada la innecesaria complejidad de la solución basada en incentivos, optará, para su explicación del paso de un estado de naturaleza a uno político, por la solución que tiene en cuenta los acuerdos de auto-interés. Estos acuerdos se constituyen gracias a la capacidad que tienen los individuos de comunicar sus preferencias por una u otra acción que favorece a los propios intereses. Al ser comunicadas, dichas preferencias pueden coordinarse con las de los demás individuos presentes en la interacción y, así, generar entre ellos confianza en que el otro cumplirá la acción acordada para el beneficio de todos. Hampton acertadamente denomina estos acuerdos como de auto-interés, puesto que en ellos prima el interés individual y el cálculo racional del que realiza la acción acordada, sobre los incentivos o el temor a un poder externo.

En este sentido, vemos que si varios individuos se encuentran discutiendo la posibilidad de un pacto para instaurar a un soberano (porque por casualidad no sufren en ese momento de *miopía*

*cognitiva*), en primer lugar, hay que suponer que lo hacen porque se han dado cuenta que para sus propios intereses es mejor llegar a un tal acuerdo, que continuar en un estado de guerra virtual. En segundo lugar, esta actitud ya es una muestra de que en principio cada uno de ellos tiene esa preferencia auto-interesada. Así que, en tercer lugar, los individuos pueden coordinar su respectivo interés por salir del estado de guerra mediante la aceptación del pacto. En este sentido, la preferencia de un individuo puede cambiar de no querer hacer un trato a querer hacerlo porque entiende que los demás también quieren constituirlo. Como sostiene Binmore (1994) “Si el contrato social incluye un acuerdo para coordinarse en un equilibrio adecuado, entonces la gente respetará los términos del contrato social porque hacerlo redundará en su propio beneficio” (370). Veamos la matriz de ganancias de una acuerdo de auto-interés o problema de coordinación.

Tabla 3.4. Matriz de ganancia, Juego de auto-interés.

		B	
		Pactar	No pactar
A	Pactar	4;4	1;3
	No pactar	3;1	2;2

Como apreciamos, la matriz de ganancias de un acuerdo de auto-interés (idéntica a la de DPI, pero por distintas razones<sup>8</sup>), expresa dos equilibrios de Nash (4;4 - 2;2) y ninguna estrategia dominante (no hay una única estrategia que, sin importar que haga la contraparte, se pueda elegir). La clave de esta ausencia de estrategia dominante tiene que ver con que a diferencia del DP, desertar mientras el otro coopera no es el mayor pago (téngase en cuenta el daño a la reputación por emplear dicha estrategia, o la pérdida de tiempo, recursos y demás, por buscar un trato necesario pero que nunca se da). Esto provoca que el individuo no vea esa opción como lo mejor que le pudiera pasar. De esta forma se evita, por tanto, la tentación a ser *Free Rider*. Más bien, la mejor opción, el mayor pago es el de que ambos cooperen. Así pues, los dos equilibrios se dan cuando las personas se enfrentan a la decisión del otro, si mi contraparte coopera (el ideal), o entiendo que tiene serias intenciones en cooperar, debo cooperar para obtener la mayor ganancia que se pueda; si mi contraparte no coopera o no tiene la intención de

<sup>8</sup> Distintas razones dado que para el caso de DPI la matriz se configura a partir de la esperanza que los individuos tiene de más ganancias en interacciones futuras, mientras que el AG o juegos de coordinación, la matriz se esboza a partir de la interdependencia entre las decisiones de los individuos, todos quieren, por su auto-interés, colaborar; hacerlo depende de si los demás colaboran o demuestran querer hacerlo.

hacerlo, no debo cooperar para no sufrir daño en el trato. Todo esto se da dentro del egoísmo psicológico y la racionalidad instrumental de los agentes hobbesianos.

Para Hampton una de las formas de romper con el problema de coordinación en donde hay dos o más equilibrios es precisamente mediante incentivos, positivos o negativos, dirigiendo la preferencia de los individuos a uno de estos equilibrios. Para el caso de los individuos hobbesianos auto-interesados, lo que hay que mostrarles es que si una buena parte de ellos están dispuestos a cooperar en la instauración de un contrato social, entonces la mayoría, sirviéndose de este conocimiento, cooperará. Esta estrategia que propone Hampton se concentra en erradicar la *miopía de segundo orden* de los individuos hobbesianos, a través de un acto que lleven a cabo estos para demostrar que van a tomar determinado curso de acción; este acto puede consistir en una promesa, por ejemplo. La promesa, en otros términos, es una herramienta eficiente para que las personas coordinen sus preferencias y con cierto grado de confianza tomen en serio la posibilidad de cooperar con aquella persona que promete. Hampton (141) afirma que en ocasiones no es necesario realizar un acto verbal para demostrar que se está dispuesto a preferir un curso de acción en lugar de otro. Así pues, la gente puede confiar en otra persona si siempre que la ha visto actuar, aquella ha respetado la reciprocidad de un acto cooperativo.

Ahora bien, creo, como parece también sugerirlo Hampton, que a partir de este primer paso hacia la construcción de confianza, se pueden ir generando pequeños grupos de individuos que cooperan entre sí, pues su elección se sustenta en la confianza mutua de sus promesas de acción. Estos grupos irán “arrinconando” o presionando a aquellas personas que aún no eligen cooperar, puesto que será menos ventajoso para un individuo competir o defenderse de un grupo de individuos coordinados. En últimas, los individuos que aún prefieren los beneficios de corto plazo dejarán de hacerlo cuando racionalmente calculen que el no estar en un grupo puede ser más peligroso, también a corto plazo, para su autopreservación. Desde otro punto de vista, los individuos que no estén en una coalición se verán forzados a unirse por el simple hecho de que tal estrategia es la más adecuada para mantener sus propios intereses, ya sea porque ven que los comunitarios se esfuerzan menos y obtienen más o porque los comunitarios representan una gran amenaza frente a sus bienes y propia vida. Así, la necesidad de elegir la estrategia de unirse o aceptar el pacto de unión bajo una misma fuerza puede, también, interpretarse como una estrategia adaptativa que elige la solución que mejor funciona

en el ambiente. Esto, visto desde una perspectiva de teoría de juegos evolutiva se puede esbozar como aparece en la siguiente matriz:

Tabla 3.5. Matriz de ganancia, Juego de Coalición.

		B	
		Coalición	Individual
A	Coalición	5;5	4;2
	Individual	2;4	3;3

En este caso, es claro que no hay ningún dilema, la estrategia dominante es cooperar, pactar la coalición (no importa que elija la contraparte, siempre será mejor elegir hacer la coalición, pues si la contraparte elige hacer coalición se obtendrá en mayor beneficio que se pueda alcanzar, pero si la contraparte no quiere hacer coalición, sigue siendo el segundo mejor pago elegir hacer la coalición); el equilibrio de Nash es la coalición y el óptimo de Pareto también. Así pues, en un estado de “semi-naturaleza” en donde algunos individuos lograron superar la *miopía cognitiva* y alcanzaron una coalición dirigida por un soberano, es altamente irracional que los demás individuos no decidan unirse a esta o constituir otra coalición.

En definitiva, es importante resaltar que la solución propuesta por Hampton (el problema de coordinación o acuerdo de auto-interés) para superar el estado de naturaleza de Hobbes es eficiente y completamente coherente con los presupuestos hobbesianos de la psicología egoísta y la racionalidad instrumental. Así pues, Hampton logra con su propuesta una explicación de la cooperación en situaciones de alto riesgo sin acudir a presupuestos distintos de los que ya Hobbes asume.

Quiero ahora, tratar con más detalle las características y diferencias entre los problemas de coordinación (a los que Hampton ha acudido) y los problemas de cooperación (estilo DP), para abordar con mejores herramientas la propuesta de Hobbes del papel del soberano absoluto en un estado político o civil. Este papel, como ya dijimos, se concentra en limitar el poder de sus súbditos, infligir castigos y ofrecer incentivos discriminadamente para eliminar la *miopía cognitiva* de las personas y, así, provocar que la cooperación sea para ellos racional y preferible.

### ***3.2 Problemas de coordinación***

Por décadas se han estudiado los problemas de cooperación bajo el paradigma de la racionalidad perfecta y que está en la esencia de juegos tan famosos como el del DP, en sus distintas versiones, o el de Bienes públicos. Así, siempre se ha pensado que el problema fundamental radica en la irracionalidad de cooperar y, por tanto, los individuos no albergan en principio esta posibilidad. Si el problema, por el contrario, no radica en que los individuos no quieren cooperar o no albergan la posibilidad (pues como lo vimos en el apartado anterior, la cooperación puede llegar a ser racional), sino que existen múltiples factores que hacen que las preferencias de la cooperación no se conozcan o no lleguen a feliz término (ya vimos el ejemplo de la miopía que propone Hampton), entonces buena parte de los problemas de la cooperación deberían transformarse en problemas sustancialmente distintos que se enmarquen bajo la denominación de problemas de coordinación. Avances en la psicología cognitiva y evolutiva han mostrado que los problemas de la cooperación pueden verse en un escenario un poco más real que incluya, no solamente el auto-interés humano, sino que a su vez se tengan en cuenta las preferencias de los demás que interactúan con nosotros. Es decir, que más que a una estrategia eminentemente optimizadora se pueda acudir a una estrategia adaptativa que evalúe las estrategias del ambiente (Ostrom, 1998; Reuben, 2003).

Bajo esta consideración, creo que si nos concentramos en explicar los problemas de la cooperación mediante la única lógica del cálculo de costos y beneficios en la contribución de individuos egoístas a objetivos colectivos, muy seguramente, como ha venido sucediendo, no podremos explicar gran parte de la cooperación que, a diferencia del laboratorio, se da en la vida real de los seres humanos. Así pues, es conveniente, sin dejar a un lado el presupuesto racionalista y auto-interesado, revisar otras consideraciones que tengan en cuenta una racionalidad acotada y la coordinación de preferencias individuales para el logro de objetivos comunales, a su vez, más beneficiosos para cada individuo (Oberschall, 1994; Moore, 1995; Hardin, 1995).

Veamos ahora, siguiendo la argumentación de Miller (2007) las características que diferencian a los problemas de coordinación (vg., Juego de Confianza -AG-) de los problemas de cooperación (vg., Dilema del Prisionero -DP-).

	<b>Problema de coordinación (AG)</b>	<b>Problema de cooperación (DP)</b>
<b>Primero</b>	Sin estrategia dominante. Pero, la cooperación es la de mayor preferencia.	Con estrategia dominante (Desertar).
<b>Segundo</b>	Para explicar la cooperación se mantienen los axiomas de la teoría de la elección racional (Janssen, 2001).	Para explicar la cooperación en ocasiones es necesario apartarse de la racionalidad individual.
<b>Tercero</b>	La desertión unilateral no es la mayor ganancia. Se gana o se pierde conjuntamente (Hardin, 1995).	La no cooperación unilateral (explotación de la cooperación de otros) entrega la mayor ganancia. Incentivo a <i>Free Riders</i> .
<b>Cuarto</b>	Explicación de la cooperación mediante la superación de problemas cognitivos ( <i>miopía</i> ). Generación de información y conocimiento común.	Explicación de la cooperación mediante elementos motivacionales.
<b>Quinto</b>	Las estrategias cooperativas se refuerzan a sí mismas. La cooperación es uno de los equilibrios de Nash que se presentan. Una vez inclinada la balanza hacia uno de los dos equilibrios, los agentes no tienen incentivos individuales	Solución inestable, dado que la cooperación no hace parte de una de las soluciones predefinidas del juego. No es un equilibrio de Nash. Por tanto, siempre existe la tentación de <i>Free Rider</i> .

	Problema de coordinación (AG)	de Problema de cooperación (DP)
	para desviarse de dicha estrategia.	

Ahora bien, resaltando las características de los problemas de coordinación, encontramos que la preferencia más relevante de los agentes que participan en un tal juego siempre es la de cooperar. Además, la cooperación se puede explicar sin renunciar a los supuestos teóricos de la elección racional. Simplemente, se debe tener en cuenta que es necesario que los agentes involucrados han de compartir información y desarrollar algún tipo de comunicación eficiente (superación de problemas cognitivos) que permita que los demás se enteren de sus preferencias, para así elegir uno de los dos equilibrios de Nash; el de la cooperación. En este mismo sentido, dado que los problemas de coordinación incluyen en su matriz, como uno de sus equilibrios, la cooperación, esta estrategia es autoreforzada, puesto que al seleccionarla los agentes no presentarán incentivos distintos para no mantenerla.

Esto es precisamente lo que explica el éxito en la propuesta de Hampton (1986), con relación al paso de un estado de naturaleza a uno civil bajo el marco del *Leviatán*. Hampton argumenta que si se asume como el principal problema para llegar a un acuerdo en el estado de naturaleza a un tipo de dificultad cognitiva en los individuos hobbesianos, la *miopía cognitiva de segundo orden*, podemos llegar a una solución mediante la facilitación de comunicación de preferencias entre los agentes. Para que dos o más individuos coordinen sus preferencias y puedan cooperar, lo único que se debe lograr es que entre ellos sepan que el otro elegirá cooperar. Si hay algún modo en el que cada uno sepa cuál es la preferencia del otro, el dilema se diluye fácilmente. Con todo, es la generación de información y conocimiento común el elemento más importante para que un problema de coordinación se resuelva.

La coordinación de preferencias también se puede lograr mediante el seguimiento a ideales, líderes u objetivos institucionales. Cada uno de estos elementos puede ser catalizador o fenómeno desencadenante (*tipping phenomenon*) (Miller, 170) del conocimiento común de las preferencias de los agentes y, por tal motivo, todos ellos funcionan como facilitadores de la cooperación (Chwe, 2001; Camerer, 2003; Miller, 2007).

Es importante, ahora, revisar con más detalle algunas soluciones a los problemas de coordinación que, por supuesto, presuponen de algún modo la comunicación y la información común. Una de las principales soluciones tiene que ver con las *promesas* conjuntas. Este tipo de acciones hace que un individuo sepa que el otro o los otros comparten las mismas preferencias que él. Sin embargo, uno de los mayores problemas con este tipo de solución directa, es que en muchos casos el ambiente en el que se desarrolla una interacción entre individuos, presenta problemas para que ellos puedan comunicarse entre sí (problemas cognitivos). Esto tiene que ver, por ejemplo, con características geográficas que aíslan a los agentes, dificultades técnicas que generan graves problemas de comunicación, etc.

De otra parte, incluso, existiendo condiciones para que la comunicación se dé efectivamente, puede ser que no sea perfecta. En este caso, se requiere de estrategias que refuercen los acuerdos y *promesas* para mantener los tratos.

Otro tipo de soluciones para problemas de coordinación se relacionan con las "*saliencias*" o *prominencias*. En este caso no son necesarias las *promesas*. Porque simplemente los individuos, en determinadas situaciones, pueden identificar que un equilibrio o solución a un juego de coordinación es más prominente que otro (Schelling, 1960; Lewis, 1969). Por su puesto que cuando esto se da, si se presenta un grado muy alto de *prominencia* o claridad con la que se expresa una solución frente a otra, nos encontraremos, entonces, con un juego de un solo equilibrio de Nash. Estas *prominencias* o "*saliencias*" pueden alimentarse de múltiples factores como por ejemplo la acumulación de experiencias pasadas en situaciones similares o con los mismos individuos, o identidad de grupo, entre otros.

El *aprendizaje adaptativo* es otra solución muy estudiada para los problemas de coordinación (Young, 1997). En este caso la solución se hace más relevante, la selección de una estrategia colaborativa puede ser más clara para los individuos, cuando encuentran que a lo largo del tiempo esa estrategia, en otros o consigo mismo, ha funcionado bien. La prueba y el error es una manera adecuada para encontrar soluciones de coordinación de preferencias.

Ahora bien, es importante anotar que a pesar de que los problemas de coordinación y de cooperación poseen características independientes, por lo tanto, son analíticamente distintos, se pueden encontrar interactuando en numerosas situaciones reales. Una, en la que me concentraré en el siguiente apartado, tiene que ver con aquellos casos en los que la coordinación genera un poder (soberano) que se instituye para desarrollar acciones coercitivas

mediante sanciones o propositivas, mediante incentivos (que cambian la estructura del DP) para consolidar la cooperación entre individuos.

### ***3.3 Sanciones e incentivos. El papel del soberano hobbesiano***

El acuerdo entre hombres libres, en medio de un estado de naturaleza peligrosa para la seguridad individual, puede lograr que un soberano se instaure como máxima autoridad llamada a pacificar las relaciones colectivas (XVII. 137/128), haciendo que los individuos salgan de un DP, o más bien que se percaten de que no están en él, y puedan, o coordinar sus preferencias con mayor facilidad, o incursionar en un juego de DPI. Este papel es desempeñado por el gobernante, dado su máximo poder (entregado por acuerdo), a través de distintos mecanismos tales como: proferir leyes civiles, amenazar a los potenciales infractores y no cooperadores, castigar la deserción frente al cumplimiento de las leyes o incentivar su seguimiento con la finalidad de hacer racional la cooperación.

En este sentido, el soberano, después de ser instaurado, tiene el poder de transformar lo que en apariencia se veía en el estado de naturaleza como un DP, en una dinámica colaborativa.

Junto con el poder de transformar estructuralmente los DP, el soberano puede hacer que se dé efectivamente la mutua realización de promesas, de tal manera que con ellas se expresen las preferencias de las partes participantes en un acuerdo y, así, se afirme la adquisición de un compromiso. Para Hampton (209), la promesa es fundamental para transformar la estructura de un DP, pues desde una perspectiva humeana el temor al castigo nos obliga a señalar, mediante una promesa, que preferimos determinado curso de acción que nos compromete con otra persona.

Así pues, Hobbes sostiene que los derechos y facultades que constituyen la soberanía en ejecución son doce (XVIII. 142/133); entre los cuales, los más relevantes son, proferir leyes civiles que regulen la vida en comunidad, junto con el diseño de incentivos y sanciones que le permitan al soberano obligar el seguimiento de los preceptos civiles instaurados por el gobierno. En este punto es importante señalar que el seguimiento de las leyes naturales, según Hobbes, es connatural a la racionalidad de cada uno. Así pues, todos los hombres deben buscar a toda costa su supervivencia, incluso si para esto se tienen que enfrentar al mismo soberano (XXI. 177/167) y, por otro lado, todo hombre debe tratar a los demás como él quisiera ser tratado (XIV. 107/100). Claro, en este último caso, prima la ley de autopreservación. De otra

parte, también es facultad fundamental del soberano impartir castigos y recompensas a quienes violen las leyes por él instauradas. Los castigos y recompensas como dispositivos “sociales” son para Hobbes, resalta Hardin (1990), de suma importancia para mantener cohesionado al Estado, ese hombre artificial (Leviatán) con el poder suficiente para hacer que los individuos se mantengan en paz.

[...] se asigna al soberano el poder de recompensar con riquezas u honores, y de castigar con penas corporales o pecuniarias, o con la ignominia, a cualquier súbdito, de acuerdo con la ley que él previamente estableció; o si no existe ley, de acuerdo con lo que el soberano considera más conducente para estimular los hombres a que sirvan al Estado, o para apartarlos de cualquier acto contrario al mismo. (XVIII. 148/138).

Hardin (1990, 79) afirma que desde el punto de vista de Hobbes, representante del contractualismo de los siglos XVII y XVIII, y teniendo en cuenta las propuestas de Locke, fundador de las teorías del derecho natural, los gobiernos fracasan si en ellos no se ejerce la coerción de los comportamientos individuales. En este sentido, el proyecto político de constituir un Estado y mantenerlo en el tiempo para consolidar la situación civil, alejada del desorden y el conflicto del hipotético estado de naturaleza, no es viable si en él (en el estado político o civil) no se presuponen elementos coercitivos que, ejecutados por el soberano, mantengan a los individuos temerosos de incumplir con el pacto de vivir en sociedad. Yo añadiría, como creo que lo asume Hobbes también, que además de los elementos coercitivos son necesarios en el mismo sentido, elementos o dispositivos de recompensa o incentivos. Hobbes sostiene que el castigo y la recompensa “vienen a ser los nervios y tendones que mueven los miembros y articulaciones de un Estado” (XXVIII. 262/246).

Ahora bien, pero ¿cuál es la justificación de la coerción y la recompensa en la teoría hobbesiana del Estado? En principio, como el mismo Hardin (1990) lo resalta, es importante señalar que la instauración de un soberano mediante el contrato social suscrito por los individuos en situación de guerra (al estilo hobbesiano), se desarrolla de manera racional y mediante un pleno consentimiento de estos. Así que este consentimiento racional debe extenderse también a las facultades y dispositivos de gobierno necesarios que tal soberano deberá ejecutar para cumplir con el objetivo por el cual ha sido elegido.

Así, en general, la mayor parte de nuestras interacciones se llevan a cabo frente a otra persona (relación diádica) y, por lo común, estas relaciones están atravesadas por intercambio de promesas, explícitas o no, que mantienen la colaboración en un punto estable<sup>9</sup>. Teniendo en cuenta esto, se presentan situaciones difíciles en las que consentimos que nuestros gobernantes asuman sistemas de coerción para ejecutar castigos, de tal forma que podamos disfrutar de los beneficios de los contratos y las promesas. Sabemos, y lo consentimos racionalmente, que los beneficios de la cooperación instaurada mediante contratos y promesas nos ofrecen mayores réditos que el no cumplir. Por esta razón aceptamos la posibilidad y la amenaza de castigo por el incumplimiento, si una de las dos partes decide desertar. Consentimos esta posibilidad a pesar de que el castigado puede ser uno mismo. Somos conscientes de los beneficios del sistema, incluso si yo resulto castigado por su ejecución. Piénsese, por ejemplos, en situaciones comerciales en donde las tarifas y los precios han sido definidas de antemano por un ente regulador (gobierno o soberano). Para mi, como cliente está bien que las tarifas de algunos servicios no se definan al gusto del prestador (podría abusar de su monopolio). De la misma forma, para el prestador no es bueno competir con tarifas que otros, con mayores capacidades y recursos, pueden ofrecer, explotando sus sistema de tarifas con precios más bajos. Ahora, ambas partes sabemos que si uno de los dos no respeta estas tarifas preestablecida, caerá sobre aquel una sanción estatal que será mejor que coopere y respete dichas tarifas. En este sentido, como lo afirma Hardin (1990, 80), no podemos concebir a las sociedades modernas sin respaldar sus relaciones contractuales con sistemas de coerción. Hobbes diría, en este sentido, que sin la espada, los pactos solo son meras palabras (XXVIII). En esto radica el consentimiento racional. Al asumir y aceptar los sistemas coercitivos, comprendiendo la superioridad de los beneficios que dichos sistemas pueden asegurar en las relaciones contractuales, actuamos de manera racional y buscamos nuestros propios intereses.

De la misma forma, en interacciones multi-agente como la provisión de bienes públicos, la influencia de los sistemas de coerción es fundamental para poder disfrutar de los beneficios de la cooperación. Casos, como los sistemas de control de tráfico en donde aparecen vinculados comparendos y sanciones para infractores, son consentidos racionalmente, pues el simple hecho de que todos contemplen la posibilidad del castigo en estas situaciones, nos permiten beneficiarnos de la buena conducción de la mayoría de las personas que están al frente del volante. Aunque siempre conducimos con precaución, pues anticipamos alguna posible

---

<sup>9</sup> Como lo vimos en secciones anteriores, las promesas cumplen un importante papel en la coordinación de preferencias.

violación de la norma, de alguna manera sentimos cierta confianza en que, en general, ese hecho constituye una probabilidad muy baja. De lo contrario, nunca conduciríamos en las calles. Si consideramos que las posibilidades de la infracción son altas, no pondríamos en riesgo nuestras vidas, ni las de nuestro acompañantes.

Todos estos elementos se siguen enmarcando en la teoría hobbesiana del auto-interés como fundamento psicológico del comportamiento humano, además del principio de autopreservación que gobierna a los hombres. De acuerdo con esto, y como acertadamente lo sostiene Hardin (1990, 80), el interés racional es suficiente para explicar los sistemas de coerción.

Desde el punto de vista de Hobbes, los súbditos admiten la coerción de un soberano, pues mediante ella se protegen de las pasiones de otros y, por qué no decirlo, de ellos mismos, al extraerlos, con esto, de una dinámica del DP. Vistas así las cosas, podría decirse que buena parte del bienestar del Estado y de las personas que en él interactúan se debe al gobierno coercitivo que lo dirige. Recuérdese que para Hobbes la vida sin gobierno es severa e improductiva; hasta peligrosa para la seguridad personal. Así pues, la justificación de la coerción en el gobierno hobbesiano se encuentra en el hecho de que mediante los castigos (de la misma forma que con los incentivos) la estructura de un juego de DP se transforma fácilmente a una de coordinación de preferencias o, también, a un de DPI.

Así, si dos o más personas se encuentran aportando para un bien público, pongamos por caso impuestos para la construcción de caminos, sin un poder coercitivo, como ya lo vimos en el capítulo 2, las preferencias de los individuos serán las de siempre desertar, dado que encuentran más rentable explotar la cooperación de los otros o, en su defecto, prefieren no cooperar para que otros no los exploten. Esta interacción tiene la típica matriz de preferencias de un DP, o de un juego de bienes públicos. Sin embargo, si existe un poder coercitivo que amenace con infringir castigos a los individuos no cooperadores y que les haga ver que explotar la cooperación puede llegar a ser menos ventajosa de lo que creían, entonces sus preferencias se transforman en actitudes cooperativas que, por supuesto, son más rentables que la deserción asumiendo el castigo. De esta manera, un sistema de castigos transforma la estructura de un juego de DP o bienes públicos a uno de coordinación de preferencias o DPI. Hobbes sostiene algo muy cercano a lo dicho, pues al menos en una ocasión del *Leviatán* (XXVII. 241/226) afirma que los castigos que no son suficientemente grandes para disuadir al infractor, lo inducirán para

violar la norma, pues este solo calculará que el beneficio de desertar es mayor que el propio castigo.

En efecto, si de antemano se conoce el castigo, cuando éste no es bastante grande para disuadir de la acción, constituye un estímulo para ella, porque cuando los hombres comparan el benéfico de la injusticia por ellos cometida con el daño que representa su castigo, por razón de naturaleza eligen lo que resulta preferible para ellos” (XXVII. 241/226).

Con todo, creo que Hobbes suscribiría un punto de vista como el sostenido anteriormente y que se refiere al poder que tienen los castigos, y de la misma forma los incentivos, en transformar un DP en un problema de coordinación de preferencias. Esto es fundamental, puesto que así se ratifica filosóficamente el importante papel del castigo y los incentivos en la estabilización de la cooperación. Sin embargo, el problema con el que nos encontramos al final del capítulo 2 no ha sido superado en la actual visión. A saber: que el sistema de castigos puede ser explotado por los *Free Rider* como un bien público de segundo orden.

Aquí me quiero detener un momento en la obvia solución que extrañamente pocos han considerado. Por lo menos, en el caso de la especie humana es evidente que un sistema de castigos pudo influir sustantivamente en la evolución de la cooperación. Hemos encontrado un seguro que protege los sistemas de castigo de la explotación de los desertores, evitando que el sistema se disuelva en una reducción *ad infinitum* de castigos de primer, segundo, tercer orden, etc., para luchar en contra de los *free riders* que aparecen en todos los niveles (*cf.*, sección 2.4).

De acuerdo con Hobbes el gobernante goza de un gran honor y dignidad por desempeñar sus funciones en el mantenimiento del orden social (XVIII. 149/140). El poder que se le entrega junto con la dignidad y el honor es un mayúsculo pago por sus servicios como dirigente del Estado y, por su puesto, como supremo ejecutor del sistema de coerción. De la misma forma, menciona Hobbes, que a aquellos quienes por su labor entregan servicios al Estado, entre los que podemos contar a los cuadros de control, como los denomina Hampton, o cuadros de coerción institucional, se les deben entregar recompensas (salarios o sueldos), en reconocimiento a las labores realizadas por contrato (XXVIII. 260/245). Es más, Hobbes sostiene que a algunos individuos se les entregaran incentivos en liberalidad (sin contrato de por medio) por el mero hecho de que con ello se incentiva el mejor servicio al Estado (XXVIII. 260/245), pongamos por caso, personas no vinculadas a los servicios del Estado que denuncian las deserciones.

En el primer caso, sostiene Hobbes, el individuo queda obligado por contrato a desempeñar las funciones del cargo, ya sean las de legislador, juez o integrante de un cuerpo de coerción. Este es el punto que quiero resaltar. Es evidente que, manteniendo los presupuestos de la psicología egoísta de Hobbes y aquellos enmarcado en la teoría de la elección racional, los cuadros de control o coercitivos deben ser recompensados con incentivos tipo salario y demás, suficientemente bien como para que los costos de desempeñar funciones de castigadores sean menores frente al beneficio obtenido, que en muchos casos también se suma a esto el honor y el poder. En este sentido, son los incentivos los que transforman la estructura de pagos de los DP de segundo orden, para generar juegos de coordinación. Tanto el soberano hobbesiano, como los individuos que llevan a cabo el “trabajo sucio” de castigar, son recompensados por el Estado, superando los costos de sus funciones, con recursos generados por la comunidad para tal fin. Nuevamente, aparece aquí el racional consentimiento de las personas para aceptar y aportar a un sistema de coerción que se auto refuerza mediante los aportes que el sistema mismo vigila para que se entreguen.

En otras palabras, los castigos institucionales derivados de sistemas coercitivos estatales rompen con los DP de ordenes superiores, puesto que mediante las recompensas instituidas como sueldos, más aquellas vinculadas a su misión como el honor y el poder, se superan los costos que deben asumir los individuos castigadores institucionales. De tal suerte que el castigo ya no es un fenómeno extraño de tipo altruista, tanto biológico como psicológico, que aporta a la estabilidad de la cooperación, y que no puede explicarse desde un punto de vista racional. Más específicamente, dado que los castigos institucionales reportan más beneficios que costos a quienes los ejecutan, dicho comportamiento no puede enmarcarse, desde un punto de vista evolutivo, bajo el altruismo biológico de ceder beneficios a los demás en contra de su propio éxito adaptativo (invertir en el fortalecimiento de un sistema de castigos para que todos se beneficien). Por otro lado, creo que este tipo de castigos, así considerados, tampoco pueden enmarcarse bajo la denominación de altruismo psicológico, pues el ejecutor del castigo institucional está motivado más por su propio interés de lograr recompensas y reconocimientos que por el mero hecho de castigar a alguien por aversión al delito o por empatía y solidaridad hacia los explotados.

Así pues, bajo las dinámicas de la elección racional, el castigo institucional en la especie humana es la solución más consistente a la estabilidad de la cooperación. Como lo hemos visto ya Hobbes lo había contemplado.

Los castigos institucionales, enmarcados en un sistema de control y coerción que se autoreforzan con la vigilancia del aporte a este sistema y a las demás empresas comunales, transforman efectivamente la estructura de pagos de un DP, o de bienes públicos en interacciones multi-agente, para convertirlos en problemas de coordinación que tiene solución en puntos de equilibrio colaborativos. Recuérdese que (*cf.*, sección 3.2) gracias a que los juegos de coordinación contemplan la cooperación como una solución posible, la estrategia que desemboca en este tipo de equilibrio es autoreforzada, en tanto que los individuos no presentan preferencias distintas a seguir cooperando.

Puede interpretarse que Hobbes, en algún sentido, consideraba los castigos y, por su puesto, los incentivos como dispositivos que coordinan preferencias. Para Hobbes los castigos establecen ejemplos para que los posibles infractores prefieran cooperar, “[...] la finalidad del castigo no es la venganza y la descarga de la ira, sino el propósito de corregir tanto al ofensor *como a los demás, estableciendo un ejemplo*”. (Sin resaltar en el original). (XXX. 286/269).

En otras palabras, y como ya se dijo en la Sección 3.2, los castigos pueden hacer que las personas se hagan más conscientes de sus preferencias y las coordinen con otros individuos. Las sanciones permiten identificar una *prominencia* o “*saliencia*” de un equilibrio colaborativo, por el que en principio la mayoría optará. Piénsese nuevamente en el caso del control del tráfico: si comprendo que las multas o castigos por conducir bajo los efectos del alcohol son sustancialmente altas, no tendremos muchos problemas en tomar un taxi y realizar una transacción con el conductor. Confiamos (esta es la coordinación de preferencias) que el conductor que nos toque no le interesa estar ebrio, tampoco le interesa violar peligrosamente las normas de tránsito, de tal forma que nos llevará a salvo a nuestro destino.

En otras palabras, es necesario tener en cuenta que en muchos casos, como lo sostiene Gilbert (2001), las sanciones no solamente ajustan el comportamiento de las personas al comportamiento del grupo, por el simple hecho que los individuos anticipan o prevén un posible castigo a la deserción; también, y tal vez más importante que lo anterior, los castigos hacen que los individuos recuerden que hacen parte de una comunidad con ciertas reglas de juego y que están comprometidos con el funcionamiento de dicha comunidad. Como lo vimos en el primer capítulo, la identidad de grupos es una de las soluciones a los dilemas sociales, una de las más efectivas maneras de fomentar la cooperación.

De otra parte, es bueno reafirmar que para el caso de los castigos institucionalizados, que en las sociedades humanas más tempranas ya hacen presencia (Lyons, 2003), no es necesario explicarlos como castigos biológicamente altruistas que dependen de un supuesto y controvertido proceso de Selección de grupos, más bien este tipo de mecanismos se explican mejor bajo la teoría de la elección racional, al encontrar que el beneficio por castigar es mucho mayor que el costo. Además, ese beneficio es asegurado por la misma labor hecha por los agentes de control frente a los aportadores al bien público.

Quiero, por último referirme a la importancia de los incentivos como dispositivos estabilizadores de la cooperación. De la misma forma que los castigos, si los incentivos son dispensados por un ente de control, como lo menciona Hobbes, los dilemas de ordenes superiores se desvirtúan de inmediato, puesto que también ellos se alimentan de aportes controlados y vigilados por los entes estatales a los que se les paga por dicha tarea. En principio, vimos como los incentivos o recompensas son fundamentales para romper con el círculo vicioso de los sistemas de control que eran explotados por *Free Rider*. Gracias a las recompensas adecuadas, los participantes de cuadros de castigo tiene el incentivo suficiente para realizar su labor sin hacer de sus acciones estrategias irracionales de altruismo. Se comprende, entonces, cómo los incentivos también transforman la estructura de DP hacia coordinaciones de intereses que fomentan la cooperación.

Quizás el vínculo entre sanciones e incentivos es más fuerte de lo que normalmente lo percibimos. No solamente ambos juegan un papel similar en la estabilización de la cooperación. No solamente ambos pueden explicarse mediante los presupuestos de la teoría de la elección racional. Además, a nivel neuronal, castigar produce en los ejecutores de la sanción efectos similares que cuando ellos mismos reciben recompensas (Nowak, 286). Según Nowak, los castigadores experimentan “agradables oleadas” de irrigación sanguínea en los centros cerebrales de la recompensa. Esto podría reflejar el correlato cerebral de experiencias relacionadas con el placer a castigar que casi todos sentimos cuando le “damos su merecido” a alguien, un tipo básico de *Shadenfreude*. Así pues, la ciencia neurobiológica puede ofrecernos pruebas de la evolución del castigo y los incentivos, como caminando hombro con hombro en medio de nuestra filogenia.

En cualquier caso, la recompensa, si la comparamos con el castigo, es un dispositivo que lleva a la cooperación de manera más creativa. Esto es así, porque en el caso de la recompensa las personas que transforman sus DP gracias a ella, en general, no solo cumplen con los acuerdos o

normas, en muchas ocasiones superan estas expectativas y generan nuevos vínculos de cooperación. La recompensa incentiva, no solamente presiona, la cooperación. En otro sentido, los castigos cumplen efectivamente con la estabilización de la cooperación, generando una coordinación de preferencias en puntos de equilibrio determinados, pero, también en general, los individuos en estos casos solo se limitan a cumplir con los tratos y tareas, sin ir más allá de lo prescrito o prometido (Nowak, 301). Es por esto que Nowak sostiene que “Reward, not necessity, is the true mother of invention” (301).

## 4. CONCLUSIONES

Como hemos visto, mi propuesta de explicación de la evolución de la cooperación se vale de los distintos esfuerzos que durante los últimos cuarenta años se han venido desarrollando bajo el marco de distintas disciplinas. Entre estas se encuentran, por un lado, la biología con su teoría de la evolución, definida inicialmente por Charles Darwin, y que predice que el mecanismo más importante del cambio y la adaptación es la selección natural, una selección que actúa en los individuos para presionar o no por cambios que se originan y replican a nivel genético. Por otro lado, encontramos la economía y la rama de las matemáticas de la que se han valido los economistas para predecir comportamientos en interacciones, la denominada teoría de juegos. También he utilizado herramientas de la psicología social para interpretar los resultados económicos y, por su puesto, todo ha sido analizado y -espero- aclarado conceptualmente por los desarrollos filosóficos que se han llevado a cabo alrededor del problema de la cooperación.

Como lo mencioné en el capítulo 2, aunque hay muchos científicos que proponen una salida más inmediata a la paradoja de la evolución de la cooperación, mediante la postulación de un mecanismo de selección a nivel de grupos, considero que no es necesario adoptar tal solución por dos razones principalmente. Primero, porque desde mi punto de vista la solución a partir de la selección de grupos explica el fenómeno, pero el mecanismo en sí tiene muchos “cabos sueltos”. Algunos de estos son por ejemplo: i) que los científicos aún no se ponen de acuerdo ni en el estatus epistemológico, ni en el estatus ontológico de este tipo de selección; ii) que a pesar de algunos desarrollos teóricos para demostrar la selección de grupos, los ejemplos de dichos procesos de selección son altamente controvertidos y, en su mayoría, se pueden explicar sin acudir a tal propuesta; iii) que la selección de grupos aún hoy carece de pruebas empíricas que la corroboren.

Otra razón por la que considero que no es necesario acudir a la selección de grupos como única vía de solución de la cooperación, tiene que ver con que, como lo he intentado mostrar, la cooperación en seres humanos puede explicarse bajo los presupuestos concomitantes de la selección individual y la teoría de la elección racional.

Así, como se ha mostrado en el capítulo 2, juegos evolutivos como el Dilema del Prisionero Iterado (DPI) presentan buenas razones para aceptar que la selección individual puede soportar la cooperación, aunque de manera inestable. Además, si se incluye la evolución de las sanciones y las recompensas, un sistema de cooperación puede llegar a alcanzar una cierta estabilidad

(estabilidad polimórfica en la mayoría de los casos). Como lo expliqué (capítulo 2), esta estabilidad polimórfica, lograda de mejor manera por la inclusión de un sistema de vigilancia, sigue siendo insuficiente para entender la cooperación extendida en seres humanos, puesto que, teniendo en cuenta las condiciones impuestas por la racionalidad individual de los agentes, también el sistema de vigilancia puede llegar a ser explotado, convirtiéndolo en un dilema de segundo orden. Sin embargo, valiéndonos de una racionalidad acotada (racionalidad imperfecta, que emplea información incompleta y que, además, está influida por las emociones, especialmente aquellas que se derivan de los castigos y las recompensas), podemos solventar los problemas de explotación de órdenes superiores del sistema de cooperación y sus sistemas de vigilancia, permitiendo de esta manera mejores posibilidades para explicar la cooperación tanto desde un punto de vista evolutivo como desde un punto de vista de racionalidad individual.

Ahora bien, creo que solucionar el problema paradójico de la evolución de la cooperación es solucionar, en últimas, el problema de la explotación de los sistemas de vigilancia que estabilizan la cooperación. En este caso mi propuesta es simple y directa. Postulo que si encontramos un tipo de sanciones y recompensas que entreguen a sus ejecutores beneficios que superen la inversión o el costo de ejecutar dichos incentivos, podemos alcanzar un sistema de cooperación autovigilante y estable. ¿Cómo se logra esto? Concentrándonos en los castigos institucionales que ya Hobbes propone como solución al dilema del contrato social y el seguimiento de leyes en un estado de civilidad. Así pues, afirmo que los incentivos institucionales no deben ser considerados como mecanismos altruistas, ni a nivel biológico, ni a nivel psicológico. En el primer caso, porque los ejecutores de incentivos institucionales (negativos o positivos) son recompensados por el *Statu quo* para que cumplan su función a cabalidad. Así, en lugar de ser un esfuerzo costoso, castigar o recompensar es más bien, “al final del día”, una labor llena de beneficios, recompensas u honores. En el segundo caso, puede pensarse de una manera más coherente con la racionalidad individual que los castigos y recompensas institucionales no son ejecutados por motivaciones relacionada con la empatía o la solidaridad. Pues, dado que el agente ejecutor recibe beneficios claros y superiores por realizar su función, este castiga o recompensa siguiendo la lógica que le señala su auto-interés.

Con un sistema de vigilancia estable, lejos del altruismo y, por tanto, de los oportunistas o *Free riders*, la cooperación entre individuos a escalas mayores se extiende a través de múltiples factores que, en últimas, lo único que provocan es que los individuos puedan coordinar eficientemente sus diversas preferencias para sacar el mejor provecho de la cooperación mutua. Como lo he mencionado, las recompensas y los castigos generan, presionan o acondicionan el

ambiente para que factores de solución de dilemas sociales como los estudiados en el primer capítulo aparezcan. Así pues, uno de los más importantes factores de solución de dilemas sociales que es generado por los incentivos, tanto negativos como positivos, tiene que ver con la modificación de la estructura del juego de DP hacia juegos del tipo DPI o Juegos de Confianza (AG), generando de esta forma la posibilidad de varios equilibrios de Nash que pueden coincidir en la cooperación mutua. Otro factor importante que es provocado por los incentivos institucionales es el denominado como identidad de grupos. Pues compartir un conjunto definido de reglas de juego, puede hacer que los individuos se sientan parte de un grupo al que hay que aportar y hacer respetar. Otro factor estudiado y que es provocado por un sistema de vigilancia se relaciona con las denominadas *prominencias* o “*salinencias*” que la posibilidad de castigos o recompensas provoca en las preferencias de los jugadores auto-interesados. En este sentido, puede suceder que el simple hecho de saber que la mayoría de las personas conocen la posibilidad de un castigo por la realización de una acción explotadora, nos hace evidente que la mejor opción para actuar es cooperar, puesto que la mayoría de la gente no se arriesgará a la sanción, como nosotros mismo no lo haremos. Por último, es importante mencionar que las emociones generadas por las sanciones y recompensas refuerzan cada uno de los factores antes mencionado, provocando que las motivaciones se conjuguen con las decisiones racionales para la cooperación.

Con todo, quiero aclarar, como lo he expuesto a través de este trabajo, que mi explicación se aparta de la consideración clásica de la economía que presupone a un hombre eminentemente calculador, con racionalidad perfecta e información completa. Mi propuesta asume la naturaleza del hombre racional, pero imperfecto que actúa influido por sus emociones y, por su puesto, por su auto-interés que acaso no es más que emoción racionalizada, pero al fin de al cabo razón en ejecución.

Por otro lado, mis conclusiones me llevan a pensar que, a diferencia de algunos autores como Nowak (2011) (quien sostiene que gracias a la cooperación hemos llegado a generar pensamiento racional, complejo y discursivo), la evolución de la cooperación, por lo menos el tipo de cooperación que empleamos los seres humanos, debe presuponer una co-evolución de otros mecanismos necesarios para su estabilización. Mecanismos como la propia razón y las emociones debieron evolucionar juntos, en paralelo, con la cooperación, para poder generar la posibilidad de una cooperación expandida en grandes grupos humanos.

Por último, no quiero concluir este texto sin mencionar que el trabajo y el tema siguen abiertos, tanto para nuevas discusiones, como para otros desarrollos. En este último caso, considero que es

importante analizar e identificar el papel que juegan las normas sociales en la estabilización de la cooperación. Esto lo sostengo, porque como lo manifesté en el primer capítulo, son las normas sociales una de las soluciones más integrales para superar los dilemas sociales. Los factores de solución que pueden generar llegan a ser aún más complejos y efectivos que los ya relacionados a los castigos y las recompensas. Además, las normas sociales presentan dentro de su estructura ciertos incentivos (negativos y positivos) que pueden generar emociones morales de gran potencia para la definición de cursos de acción; entre otros muchos factores relacionados con modificación de la estructura de pagos, ambientes informados, identidad de grupos, “*saliencias*” e interiorización del respeto a la cooperación.

## 5. BIBLIOGRAFÍA

Alexander, R. D. *The biology of moral systems*. Transaction Publishers, 1987.

Andrade, L. E. & Fajardo, D. “Niveles de selección y la cuestión ontológica acerca de las jerarquías biológicas”, *El Hombre y la Máquina*, 30 (2008): 86-99.

Axelrod, A. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton, NJ: Princeton University Press, 1997.

Axelrod, R. *The evolution of cooperation*. New York: Basic Books, 1984.

Axelrod, R. & Hamilton, W. D. “The Evolution of Cooperation”. *Science*, 211(4489) (1981): 1390-1396.

Bendor, J. & Mookherjee, D. “Institucional Structure and the Logic of Ongoing Collective Action”, *American Political Science Review*, 81 (1987): 129-154.

Biel, A. “The conceptual Framework of Social Dilemmas”. *New Issues and Paradigms in Research on social Dilemmas*. Eds. Anders Biel, Daniel Eek, Tommy Gärling, Mathias Gustafsson. New York: Springer, 2008.

Bicchieri, C. *The grammar of society : the nature and dynamics of social norms*. New York : Cambridge University Press, 2006.

Bixenstein, V. E., Levitt, C. A. & Wilson, K. R. “Collaboration among six person in a prisoner’s dilemma game”. *Journal of Conflict Resolution*, 10 (1966): 488-496.

Boehm, C. “Egalitarian Behavior and Reverse Dominance Hierarchy”, *Current Anthropology*, 34 (3) (1993): 227-254.

Bonacich, P., Shure, G., Kahan, J. & Meeker, R. “Cooperation and group size in the person prisoner’s dilemma”. *Journal of Conflict Resolution*, 20 (1976): 687-706.

Bowles, S. & Gintis, H. "The Origins of human cooperation", *The Genetic and Cultural Origins of Cooperation*, Ed, Hammerstein, P. Cambridge: MIT Press, 2003.

Boyd, R. "The evolution of reciprocity in sizable groups", *Journal Theory of Biology*, 132 (1988): 337-356.

Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. "The Evolution of Altruistic Punishment". Gintis, H. (Ed). *Moral sentiments and material interests: The foundations of cooperation in economic life*, MIT Press, 2005.

Boyd, R. & Richerson, P. J. "Punishment allows the evolution of cooperation (or anything else) in sizable groups". *Ethology and Sociobiology*, 13 (1992): 171-195.

\_\_\_\_. "The evolution of reciprocity in sizable groups", *Journal Theory of Biology*, 132 (1989): 337-356.

Brechner, K. C. "An experimental analysis of social traps". *Journal of Experimental Social Psychology*, 13 (1977): 552-564.

Brewer, M.B. & Kramer, R. M. "Choice behavior in social dilemmas: effect of social identity, group size, and decision framing". *Journal of Personality and Social Psychology*, 50 (1986):543-549.

Camerer, C. F. "Coordination". C. F. Camerer, *Behavioral Game Theory*, Princeton: Princeton University Press, (2003). pp. 337-407.

Cárdenas, J. C. *Dilemas de lo colectivo: instituciones, pobreza y cooperación en el manejo local de los recursos de uso común*. Bogotá: Universidad de Los Andes, 2009.

Chwe, M. *Rational Rituals: Culture, Coordination, and Common Knowledge*. Princeton: Princeton University Press, 2001.

Coleman, J. S. *Foundations of social theory*. Cambridge. MA: Harvard University Press, 1990.

Darwin, Ch. *The descent of man (1871)*. New York: Penguin Books, 2004.

\_\_\_\_\_. *El origen del Hombre (1871)*. Bogotá: Panamericana, 1994.

\_\_\_\_\_. *The origin of species by means of natural selection: or, the preservation of favored races in the struggle for life (1859)*. Ed. J. W. Burrow. London: Penguin Books, 1968.

Dawes, R. "Social dilemmas, economic self-interest, and evolutionary theory". *Frontiers of Mathematical Psychology: Essays in Honor of Clyde Coombs*, Ed. D.R. Brown, J.E.K. Smith. New York: Springer-Verlag, 1991.

\_\_\_\_\_. "Social Dilemmas" *Annual Review of Psychology*, 31 (1980): 169-193. DOI: 10.1146/annurev.ps.31.020180.001125

Dawes, R. M., McTavish, J. & Shaklee, H. "Behavior, communication and assumption about other people's behavior in a common dilemma situation". *Journal of Personality and Social Psychology*, 35 (1977): 1-11.

Dawes, R. M. & Messick D. M. "Social Dilemmas". *International Journal of Psychology*, 35(2) (2000):111-116. DOI: 10.1080/002075900399402

Dawes, R. M. & Thaler, R. H. "Anomalies: Cooperation", *The Journal of Economic Perspectives*, 2(3) (1988): 187-197.

Dawkins, R. *The Extended Phenotype: The Long Reach of the Gene*, New York: Oxford University Press, 1989.

\_\_\_\_\_. *The Selfish Gene*, New York: Oxford University Press, 1976.

Edney, J. J. & Harper, C. S. "The effect of information in a resource management problem: A social trap analog". *Human Ecology*. 6 (1978): 387-395.

Eek, D. & Gärling, T. "A New Look at the Theory of Social Value Orientation: Prosocials Neither Maximize Joint Outcome nor Minimize Outcome Differences but Prefer Equal Outcomes". *New*

*Issues and Paradigms in Research on social Dilemmas*, Eds. Anders Biel, Daniel Eek, Tommy Gärling, Mathias Gustafsson. New York: Springer, 2008.

Elster, J. *Nuts and bolts for the social sciences*, Cambridge: Cambridge University Press, 1989a.

\_\_\_\_\_. *The cement of society: A survey of social order*, Cambridge: Cambridge University Press, 1989b.

\_\_\_\_\_. *Juicios salomónicos: las limitaciones de la racionalidad como principio de decisión*, Barcelona: Gedisa, 1991a.

\_\_\_\_\_. *Tuercas y tornillos: una Introducción a los conceptos básicos de las ciencias sociales*, Barcelona: Gedisa, 1991b.

Falk, A. & Fischbacher, U. "A theory of reciprocity". *Games and Economic Behavior*, 54(2) (2006): 293-315.

Fehr, E. & Fischbacher, U. "The economics of strong reciprocity." *Moral Sentiments and Material Interests. The Foundations for Cooperation in Economic Life*, (2005): 151-193.

\_\_\_\_\_. "Third Party Punishment and Social Norms", *Evolution & Human Behavior*, 25 (2004): 63-87.

\_\_\_\_\_. "The Nature of Human Altruism" *Nature*, 425 (2003): 785-791.

Fehr, E. & Gächter, S. "Altruistic punishment in humans". *Nature*, 415 (2002):137-140.

Fehr, E. & Henrich, J. "Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism" *Genetic and Cultural Evolution of Cooperation*, Ed. Hammerstein, P. Cambridge: MIT Press, 2003.

Fox, J. & Guyer, M. "Group size and others' strategy in an n-person game". *Journal of Conflict Resolution*, 2 (1977): 323-338.

- Frank, R. H. *Passions within reason: The strategic role of the emotions*, WW Norton & Co, 1988.
- Ghiselin, M. T. *The Economy of Nature and the Evolution of Sex*, Berkley: University of California Press, 1974.
- Gifford, R. "Toward a Comprehensive Model of Social Dilemmas" *New Issues and Paradigms in Research on social Dilemmas*, Eds. Anders Biel, Daniel Eek, Tommy Gärling, Mathias Gustafsson. New York: Springer, 2008.
- Gilbert, M. "Collective Preferences, Obligations, and Rational Choice", *Economics and Philosophy*, 17 (2001): 109-119.
- Gintis, H. (Ed). *Moral sentiments and material interests: The foundations of cooperation in economic life*, MIT Press, 2005.
- \_\_\_\_\_. "Strong reciprocity and human sociality." *Journal of Theoretical Biology*, 206(2) (2000): 169-179.
- Gould, S. J. *The structure of evolutionary theory*, Cambridge: The Beknap Press, 2002.
- Hambuerger, H., Guyer, M. & Fox, J. "Group size and cooperation". *Journal of Conflict Resolution*, 19 (1975): 503-531.
- Hamilton, W. D. "The genetical evolution of social behavior. I." *Journal of theoretical biology*, 7.1 (1964): 1-16.
- \_\_\_\_\_. "The evolution of altruistic behavior." *American naturalist*, (1963): 354-356.
- Hampton, J. *Hobbes and the Social Contract Tradition*, Cambridge University Press, 1986.
- Hardin, G. "The tragedy of the commons". *Science*, 162 (1968): 1243-1248.
- Hardin, R. *One for All: the logic of group conflict*, Princeton: Princeton University Press, 1995.

\_\_\_\_\_. “Racionallity jutifying Political Coercion”. *Journal of Philosophical Research*. 15 (1990): 79-91.

Hauert, C., Traulsen, A., Brandt, H., Nowak, M. A. & Sigmund, K. “Via Freedom to Coercion: The Emergence of Costly Punishment”. *Science*, 316(5833) (2007): 1905-1907. DOI: 10.1126/science.1141588.

Henrich, J. & Boyd, R. “Why people punish defector”<sup>6</sup>, *Journal Theory of Biology*, 208 (2001): 79-89.

Hernich, J., Mcelreath, R., Barr, A., Ensminger, J. “Costly punishment across human societies”, *Science*, 312 (2006): 1767-1770.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Hill, K., Gil-White, F., Gurven, M., Marlowe, F., Patton, J. Q., Smith, N. & Tracer, D. “Economic Man' in Cross-cultural Perspective: Behavioral Experiments in 15 Small-scale Societies”, *Behavioral and Brain Sciences*, 28 (2005): 795-855.

Hirshleifer, D. & Rasmusen, E. “Cooperation in a Repeated Prisoner’s Dilemma with Ostracism”, *Journal of Economic Behavior and Organization*, 12 (1989): 87-106.

Hobbes, T. *Leviatán (1651)*. México: Fondo de Cultura Económica, 1980.

\_\_\_\_\_. *Hobbes’s Leviathan (1651)*. Oxford: Oxford at The Clarendon Press, 1909.

Janssen, M. “On the Principle of Coordination”. *Economics and Philosophy*, 17(2) (2001): 221-234.

Jerdee, T. H. & Rosen B. “Effect of opportunity to communicate and visibility of individual decisions on behavior in the common interest”. *Journal Apple Psychology*, 59 (1974): 712-716.

Joshi, N. “Evolution of cooperation by reciprocation within structured demes”, *Journal of Genetics*, 66 (1987): 69-84.

Kahneman, D. & Tversky, A. *Choices, Values, and Frames*, Cambridge: Cambridge University Press, 2000.

Kavka, G. S. *Hobbesian Moral and Political Theory*, Princeton: Princeton University Press, 1986.

Kollock, P. "Social Dilemmas: The Anatomy of Cooperation". *Annual Review of Sociology*, 24 (1998): 183-214. DOI: 10.1146/annurev.soc.24.1.183

Komorita, S. S. & Lapworth, C. W. "Cooperative choice among individuals versus groups in an n-person dilemma situation". *Journal of Personality and Social Psychology*, 42 (1982): 487-496.

Kuhlman, D. M. & Marshello, A. F. J. "Individual differences in game motivation as moderator of preprogrammed strategy effects in prisoner's dilemma". *Journal of Personality and Social Psychology*, 32(5) (1975): 922-31.

Lack, D. *Population Studies of Birds*, Oxford: Oxford University Press, 1966.

Lewis, D. *Conventions. A Philosophical Study*. Oxford: Basil Blackwell, 1969.

Lewontin, R. "The units of selection", *Annual Review of Ecology and Systematics*, 1 (1970): 1-16.

Liebrand, W. B. G. "The effect of social motives, communication and groups size on behavior and n-person multistage mixed-motif game". *European Journal of Social Psychology*, 14 (1984): 238-264.

Locke, John, and Carlos Mellizo. *Segundo tratado sobre el gobierno civil: un ensayo acerca del verdadero origen, alcance y fin del gobierno civil*. Madrid: Alianza Editorial, 2000.

Lyons, L. *The History of Punishment*. London: Amber Books, 2003.

Marwell, G. & Ames, R. E. "Experiments on the provision of public goods I: resources, interest, group size, and the free rider problem". *American Journal of Sociology*, 84 (1979): 1335-1360.

Marwell, G. & Schmidt, D. R. "Cooperation in a Three-person prisoner's dilemma". *Journal of Personality and Social Psychology*, 31 (1972): 376-383.

Mashima, R. & Takahashi, N. "The emergence of Generalized Exchange by Indirect Reciprocity". *New Issues and Paradigms in Research on social Dilemmas*. Eds. Anders Biel, Daniel Eek, Tommy Gärling, Mathias Gustafsson. New York: Springer, 2008.

Maynard Smith, J. "Group Selection" *Quarterly Review of Biology*, 51(2) (1976): 277-283.

McClintock, C. G. & Liebrand, W. B. G. "Role of interdependence structure, individual value orientation, and other's strategy in social decision making: a transformational analysis". *Journal of Personality and Social Psychology*, 55(3) (1988): 396-409.

Messick, D. M. "Alternative logic for decision making in social situations". *Journal of Economics Behavior and Organization*, 39 (1999): 11-28.

Messick, D. M. & Brewer, M. B. "Solving social dilemmas". *Review of Personality and Social Psychology*, 4 (1983): 11-44.

Miller, L. M. "Coordinación y acción colectiva". *Revista Internacional de Sociología*, 46 (2007): 161-183.

Moore, W. H. "Rational Rebels: Overcoming the Free-Rider Problem". *Political Research Quarterly*, 48 (1995): 417-454.

Nesse, R. "Why is group selection such a problem?", *Behavioral and Brain Sciences*, 17(4) (1994): 633-634.

Nowak, M. A. & Highfield, R. *SuperCooperators: altruism, evolution, and why we need each other to succeed*. New York: Free Press, 2011.

Oliver, P. "Rewards and punishment as selective incentives for collective action: theoretical investigations". *American Journal of Sociology*, 85 (1980): 1356-1375.

Olson, M. *The Logic of Collective Action: public goods and the theory of groups*. Cambridge: Harvard University Press, 1965.

Orbell, J. M., Dawes, R. van de Kragt A. "The limits of multilateral promising". *Ethics*, 100 (1990): 616-27.

Oberschall, A. R. "Rational Choice in Collective Protest". *Rationality and Society*, 6(1) (1994): 79-100.

Ostrom, E. "A Behavioral Approach to the Rational Choice Theory of collective Action". *American Political Science Review*, 92(1) (1998): 1-22.

Platt, J. "Social Traps". *American Psychologist Journal*, 28 (1973): 641-51.

Rapaport, A., Chammah, A., Dwyer, J. & Gyr, J. "Three-person non-zero-sum nonnegotiable games". *Behavior Science*, 7(1962): 30-58.

Reuben, E. *The Evolution of Theories of Collective Action*. Amsterdam: Universidad de Amsterdam, 2003.

Rosas, A. "¿Cómo evolucionaron las preferencias sociales?". *Revista de Filosofía Aurora*, 25(36) (2013): 343 - 359.

\_\_\_\_\_. "Ética Evolucionista: o enfoque adaptacionista da cooperacao humana" *Filosofia da Biologia*. Sao Paulo: Artmed, 2011.

\_\_\_\_\_. "The return of reciprocity: a psychological approach to the evolution of cooperation". *Biology & Philosophy*, 23(4) (2008): 555-566.

\_\_\_\_\_. "Las emociones morales como adaptaciones para la cooperación en dilemas sociales". *Ludus Vitalis*, 15(28) (2007): 97-118.

\_\_\_\_\_. "La moral y sus sombras: la racionalidad instrumental y la evolución de las normas de

equidad". *Crítica: Revista Hispanoamericana de Filosofía*, 37(110) (2005):79-104.

\_\_\_\_\_. (Ed). *Filosofía Darwinismo y Evolución*. Bogotá: Unibiblos, 2007.

\_\_\_\_\_. "El entorno ancestral de las normas de equidad" *Filosofía, Darwinismo y Evolución*. Bogotá: Unibiblos, 2007.

\_\_\_\_\_. "La Evolución de la Cooperación", *Estudios de Filosofía Política*. Bogotá: Editorial Universidad Nacional De Colombia, 2004.

Schelling, T. *The Strategy of Conflict*. Cambridge: Harvard University Press, 1960.

Sethi, R. & Somanathan, E. "The evolution of the social norms in common property resource use", *The American Economic Review*, 86 (4) (1996): 766-788.

Smith, A. *The theory of moral sentiments (1759)*. New York: Penguin, 2010.

\_\_\_\_\_. *An inquiry into the nature and causes of the wealth of nations: a selected edition (1776)*. New York: Oxford University Press, 2008.

\_\_\_\_\_. *Teoría de los sentimientos morales (1759)*. Madrid:Alianza Editores, 2004.

\_\_\_\_\_. *Investigación sobre la naturaleza y causas de la riqueza de las naciones (1776)*. México: Fondo de Cultura Económica, 1958.

Sober, E. & Wilson, S. D. *Unto Others. The Evolution and Psychology of Unselfish Behavior*. Cambridge: Harvard University Press, 1998.

Sripada, C. S. "Punishment and the strategic structure of moral systems", *Biology and Philosophy*, 20 (2005): 767-789.

Sterelny, K. *The evolution of agency and other essays*, Cambridge: Cambridge University Press, 2001.

\_\_\_\_\_. *Thought in a hostile world*, New York: Blackwell, 2003.

Sterelny, K. & Griffiths, P. E. *Sex and death*, Chicago: The University of Chicago Press, 1999.

Taylor Ch. *La ética de la autenticidad*. Barcelona: Paidós, 1994.

Trivers, R. L. "The evolution of reciprocal altruism". *Quarterly Review of Biology*, (1971): 35-57.

Williams, G. *Adaptation and natural selection*, New Jersey: Princeton University Press, 1966.

Williams, G. *Natural Selection: domains, levels, and challenges*, Oxford: Oxford University Press, 1992.

Wynne-Edwards, V. C. *Animal dispersión in relation to social behavior*, London: Oliver & Boyd, 1962.

Yamagishi, T. "The provision and sanctioning system as a public good". *Journal of personality and social Psychology*, 51 (1986): 110-116.

Yamahashi, T. & Takajashi, N. "Evolution of Norms without Metanorms" *Social Dilemmas and Cooperation*, Ed, Schulz, U., Albers, W. & Muller, U. Berlin: Springer-Verlag, 1994.

Young, J. *Economics as a moral science: the political economy of A. Smith*, Cheltenham, E: Elgar. 1997.