UNIVERSIDAD TORCUATO DI TELLA Escuela de DERECHO

Revista Argentina de Teoría Jurídica
Vol. 3, N° 1 (Noviembre de 2001)

LA CONFIANZA, LA RESOLUCIÓN DE CONTROVERSIAS, Y EL PROBLEMA DEL *QUIS CUSTODIET*

Geoffrey Brennan[*]

1. ¿Quis custodiet ipsos custodes?

De algún modo, esta conocida pregunta clásica– ¿quién vigilará a quienes nos vigilan?es *la* pregunta normativa central en el análisis de la teoría de la elección pública sobre la política y más en general sobre el diseño institucional. Esto es así porque la teoría de la elección pública normalmente hace lo que muchos considerarían suposiciones extremas acerca de los motivos de aquellos que poseen poder político –a saber, que quienes poseen poder político invariablemente tenderán a explotar ese poder para alcanzar sus propios fines a expensas de los ciudadanos en general-. La suposición es bellamente descripta por David Hume en una oración a menudo citada en los círculos de la teoría de la elección pública:

¨ al idear [...] la constitución, se debería presumir que todo hombre es un bribón que no tiene ningún otro propósito en toda su acción más que el autointerés¨. [Hume, ¨Of the Independency of Parliament¨, *Essays Moral, Political and Literacy* p.117-118]

[*] Programa de Teoría Social y Política, Research School of Social Sciences, Australian National University. La traducción fue realizada por Emiliano Marambio Catán, supervisada por Agustín Waisman, y corregida por Guido Pincione.

En realidad, es bastante dudoso si esta oración representa la opinión de Hume respecto de la naturaleza humana en general, o respecto de las suposiciones sobre la naturaleza humana relevantes para el análisis político en particular. Pero los expertos en teoría de la elección pública no pretenden ser exégetas: simplemente escogen la declaración de Hume porque ésta representa, con cierta elegancia del siglo dieciocho, la posición que ellos tienen.

Tal como Dennis Mueller (1989) lo expresa: ¨...el postulado básico sobre la conducta en la teoría de la elección pública, tal como en la economía, es que el hombre es un ser racional y egoísta que maximiza utilidad¨. [p. 2]

Bajo este postulado sobre la conducta [más precisamente un ¨postulado motivacional¨], es claro que la del *quis custodiet* es una pregunta seria. No será un problema explicar por qué una comunidad necesitará tener reglas e instituciones para la ejecución de esas reglas: el egoísmo racional de los agentes individuales, bajo una variedad de dificultades de dilema del prisionero, llevará a cada uno a actuar de maneras distintas a aquellas que todos preferirían. En principio, por lo tanto, existirá un argumento en favor de reglas que asegurarán la conducta apropiada en tales casos. Pero la existencia misma del problema nos alerta sobre la dificultad de encontrar una solución. Si esas reglas son ejecutadas por agentes que no son diferentes de aquellos cuya conducta la ejecución debe regular, pues entonces ¿quién vigilará a quienes nos vigilan? ¿Por qué no usarán los ejecutores sus poderes (necesariamente) discrecionales simplemente para promover sus propios intereses privados a expensas de todos los demás? Cuando todo hombre es un bribón, ¿cómo es posible escapar de una sociedad bribona?

En este sentido, la pregunta, *quis custodiet ipsos custodes*, es vista por los expertos en teoría de la elección pública como una clase de presunto desafío demoledor a cualquier forma de estructurar el

modelo que implique el ejercicio de poder delegado. Y esto es así de manera bastante explícita. La teoría de la elección pública surgió como una contra al así llamado modelo de gobierno del ¨déspota benévolo¨ que habitaba la economía política –y más ampliamente la teoría del economista sobre el estado– tal como fue desarrollado a lo largo de las décadas del cuarenta, cincuenta y sesenta. La práctica de aplicar criterios normativos a parámetros de política directamente –como, por ejemplo, en el diseño de ideales para el afinamiento de la macroeconomía; o en combatir ejemplos de ¨fallas de mercado¨ en relación con la provisión de bienes públicos; o en alcanzar justicia distributiva (de cualquier modo en que fuera entendida exactamente)- parecía presuponer que el consejero político / economista estaba ofreciendo su consejo a un ¨déspota benévolo¨.

Ambos aspectos de esta concepción, el despotismo y la benevolencia, fueron considerados objetables por los eruditos de la teoría de la elección pública –el despotismo porque quien ponía en práctica las políticas era considerado totalmente ajeno a cualquier restricción política, y la benevolencia debido a que se suponía que quien ponía en práctica las políticas estaba motivado solamente por la persecución del ¨interés público¨, tal como éste era articulado en los criterios normativos tan meditadamente provistos por el economista / consejero.

Fue específicamente desde la perspectiva del *quis custodiet* que el supuesto de benevolencia fue considerado tan odioso**.** La suposición de benevolencia no sólo hacía a un lado, de un plumazo, toda preocupación sobre el *quis custodiet* –después de todo, ¿cuál es el problema, si los vigiladores son totalmente benevolentes?- pero la suposición también implicaba que cualquier intervención en la forma de competencia electoral u otra restricción constitucional sólo serviría para evitar que los hacedores de políticas, quienes de otro modo no estarían sujetos a restricciones, hicieran el bien que de otro modo harían. En este sentido, la posición del déspota benévolo parecía totalmente *anti*democrática: simplemente no dejaba espacio normativo en el que las restricciones políticas pudieran actuar.

Asimismo, para los economistas, había un argumento totalmente incontestable contra la suposición de benevolencia –a saber, el argumento de la simetría motivacional-. Si la actuaban de un modo enteramente autointeresado, parecería irremediablemente parcial proclamar ideológicamente el ¨éxito político¨ sobre la base de agentes que actúan de un modo exclusivamente benevolente. Lo que hacía falta, observaban los críticos desde la perspectiva de la teoría de la elección pública, era un análisis de los procesos políticos democráticos que fuese totalmente consistente con el modelo de mercados que permitía el diagnóstico de la falla del mercado – el requisito de que fuera ¨totalmente consistente¨ importaba adoptar los mismos puntos de referencia normativos y las mismas suposiciones motivacionales que son aplicados normalmente en el análisis microeconómico de los mercados-. En cuanto a todo lo demás, la teoría de la elección pública era un oponente intransigente de cualquier suposición de asimetría motivacional: debía haber una insistencia en el viejo dicho popular de que no debe verse a los políticos –ni tampoco a los burócratas y jueces- como mejores ni peores que el resto de nosotros.

Por supuesto, el principio de simetría motivacional podía ser satisfecho –tal como el requisito de racionalidad – sin que esto implicara egoísmo. Los agentes tanto en los papeles políticos como en los de mercado podían ser modelados como *algo* benévolos; *algo* interesados en la sociedad. Además, tal estructura motivacional bien podría parecer manifiestamente más realista que el extremo del egoísta puro. Pero lo que aún estaría excluido es cualquier argumento normativo en favor del ejercicio del poder de gobernar que se base en la suposición de que los gobernantes son de cualquier modo moralmente superiores a, o más benévolos que, los gobernados. Si todos los argumentos normativos en favor del poder delegado dependen de alguna suposición similar es una pregunta abierta. ¿Representa el desafío del *quis custodiet*, bajo estas suposiciones motivacionales más moderadas, la clase de desafío demoledor al poder delegado que representa en el caso del abordar.

Antes de comenzar un análisis más detallado de esta pregunta, vale la pena notar que el principio de simetría motivacional, si bien es a primera vista suficientemente razonable, no está totalmente libre de objeciones. Nótese, en particular, que excluye la posibilidad de una *selección* efectiva de funcionarios públicos sobre la base de su ¨aptitud¨ para la función pública. Supóngase, por ejemplo, que los agentes no son idénticos –que algunos están más motivados por el interés público, son más benévolos, o más cumplidores por naturaleza, que otros-. Entonces podríamos pensar en la ¨virtud cívica¨ que estos agentes más benevolentes/cumplidores poseen como en un bien que tiene un valor social particularmente alto en esos empleos donde el poder delegado debe ser ejercido. Podríamos imaginar que los procesos de selección podrían ser ideados para asistir tanto a la identificación de tales personas como a su asignación, una vez identificados, a papeles sociales adecuados. Tal vez en este espíritu, por ejemplo, podríamos concebir a los procesos electorales democráticos menos como un modo de proveer incentivos a los candidatos para que ofrezcan políticas en el interés de los ciudadanos (la línea argumental habitual de la teoría de la elección pública) y más como un medio para escoger para la función pública a aquellos con un sentido fuerte del deber público. [Nótese que el principio de simetría motivacional excluiría esta posibilidad]. En la forma extrema, ilustrada en la cita de Hume, cualquier forma de selección de acuerdo con diferencias en la virtud cívica es excluida porque no hay heterogeneidad motivacional: ¨debería presumirse que *todo* hombre es un bribón¨. Nótese también que la heterogeneidad motivacional en sí misma no es suficiente. También necesitamos tener a mano dispositivos de selección adecuadamente robustos para distinguir los buenos muchachos de los malos. Y este no es un desafío menor.

Después de todo, las personas egoístas querrán ocupar posiciones de autoridad delegada precisamente por las mismas razones por las que queremos que las personas especialmente interesadas en el interés público las ocupen –a saber, que esas posiciones proveen la oportunidad para la explotación de otros-. Los egoístas racionales siempre tendrán un incentivo para hacerse pasar por personas interesadas

en la sociedad para ser designados en posiciones donde el poder discrecional asignado puede ser explotado para sus propios fines.

El principio de simetría motivacional comprende, entonces, o bien la suposición de que todas las personas son motivacionalmente idénticas o la de que, a pesar de la heterogeneidad motivacional, no es posible idear procedimientos de selección exitosos. Por cierto, la primera de estas posibilidades es altamente inverosímil. Sabemos, por ejemplo, a raíz de la evidencia creciente producida por una muy amplia variedad de experimentos de laboratorio [véase Sally (1995), por ejemplo] que los individuos difieren en su grado de espíritu cívico/benevolencia –que, por ejemplo, entre un tercio y un medio de las personas actúan ¨cooperativamente¨ en los dilemas sociales (experimentos de provisión de bienes públicos y otros similares)- y que en los juegos de división simple diferentes ¨dictadores¨ distribuyen un premio dado en proporciones diferentes entre ellos y el pretendiente relevante. El mensaje simple que parece emerger clara y robustamente de toda esta variedad de experimentos es que los agentes no son totalmente egoístas, y que algunos son bastante menos egoístas que otros.

No obstante, podríamos querer conservar la suposición de simetría motivacional ¨en promedio¨. Podríamos reconocer la heterogeneidad moral entre los agentes pero pensar que idear mecanismos de selección confiables que distingan a las personas buenas de las malas es un sueño sin esperanzas. Esto, al menos, es lo que supondré aquí. A través del argumento que sigue, asumiré que hay heterogeneidad motivacional pero que los agentes no pueden parece poco prometedora, el principio de simetría motivacional no impide las instituciones de poder delegado –al menos bajo ciertas restricciones que intentaré explicar claramente-. En otras palabras, el desafío del *quis custodiet* no es incontestable. El desafío no prueba por sí mismo y sin argumento posterior que los *custodes* no pueden hacer bien alguno. En ese sentido, los casos en los que todos son totalmente benévolos o todos son totalmente egoístas son igualmente engañosos. En ese mismo sentido, el procedimiento analítico humeano para el análisis

institucional –el procedimiento que la ortodoxia de la teoría de la elección pública rutinariamente adopta [y que he defendido previamente –por ejemplo, en Brennan & Buchannan (1985) ch. 4-] es indebidamente restrictivo.

Podría ser útil en este punto contraponer a la cita de Hume una cita de otra figura de autoridad –una cita que capture exactamente la posición acerca de la motivación que yo creo que de hecho es correcta y además adecuada en el contexto del análisis institucional-. La figura de autoridad en cuestión es Alexander Hamilton y la observación relevante viene de *El Federalista* Nº 76. Hamilton observa que: ¨... la suposición de venalidad universal es un error tan grande en el razonamiento político como la suposición de rectitud universal. Hay una porción de honor entre la humanidad que puede ser la base de nuestra esperanza.¨

En este trabajo [tanto como en otro trabajo reciente –tal como Brennan & Hamlin (2000)-] quiero argumentar exactamente a la par de estas líneas hamiltoneanas.

2. Confianza y contrato

Desarrollaré el argumento general que es presentado en términos del análisis de la confianza del ¨agente racional¨. Quiero analizar específicamente las implicaciones –si es contratos celebrados y/o las promesas hechas. El análisis está propuesto no tanto como una obra de análisis económico del derecho sino como una obra de teoría política. Por lo tanto, no me preocuparé por el contenido preciso de las leyes. Tampoco me preocuparé por modelar la estructura institucional de los tribunales del modo más plausible. De hecho**,** la imagen de los tribunales que ofreceré es tan lejana de cualquier institución legal prevaleciente que prefiero hablar de instituciones que ¨resuelven controversias¨, como en mi título, en lugar de hablar de tribunales como tales. Sin embargo, las instituciones que

resuelven controversias son entendidas de manera tal que una vez que el juicio toma lugar la decisión del juzgador es de cumplimiento obligatorio: los *custodes* en la explicación que desarrollaré sí tienen poderes genuinos. Las suposiciones particulares que he hecho sobre los tribunales están diseñadas para dar cabida fácilmente al principio de simetría motivacional –no necesariamente para reflejar la realidad-.

Mi objeto principal en esta sección es presentar las dificultades sociales que la economía de la confianza expone, y discutir brevemente ciertas ¨soluciones¨ posibles para esa dificultad. Todo el tratamiento aquí será lógico e ilustrativo. Las conclusiones sí derivan de un tratamiento más técnico [Brennan, Guth, & Klien (1992)] pero el análisis más elaborado no es necesario ni decisivo para los puntos centrales. Mi objeto aquí es argüir en favor de esas conclusiones de manera intuitiva y en un modo que se dirige más explícitamente a las implicaciones de teoría política.

El análisis de la confianza se ha convertido recientemente en algo así como una pequeña industria dentro de las ciencias sociales académicas. Los últimos años han visto varios libros e incontables artículos sobre el tema y sobre preguntas relacionadas acerca de la erosión y el mantenimiento del ¨capital social¨. Los economistas no han estado exentos la sociedad comercial: la confianza hace que las ruedas giren más suavemente (baja los costos de transacción) y hace posibles ciertos intercambios mutuamente beneficiosos que de otro modo no ocurrirían. En verdad, podría argüirse que algún grado (posiblemente pequeño) de confianza es necesario en *todas* las transacciones económicas: alguien siempre debe mover primero en cualquier transacción económica y así ¨confiar¨ en que quien mueve segundo cumplirá con el pacto.

Para el economista, el análisis de la confianza y de la confiabilidad debe ser encajado dentro de un sistema de racionalidad amplia –aunque como veremos, de algunos modos el fenómeno de la confianza representa cierto desafío a aquel sistema-.

En la explicación económica habitual de la confianza, toda acción de la parte que confía es impulsada por la racionalidad: es en el lado de la confiabilidad de la confianza dentro de las relaciones de confianza que el desafío a la racionalidad aparece, y en este aspecto de la confiabilidad que la mayoría del interés intelectual se enfoca.

Considérese, específicamente, el juego básico de confianza, designado Fig. 1. Es un juego secuencial de dos personas, representado aquí en forma extensiva. El jugador I mueve primero y debe elegir entre no confiar (N) y confiar (C). Si I elige N, el juego termina y ambos jugadores reciben un pago de cero: los pagos están representados en cada nodo por un par de números (*a, b*) donde *a* es el pago para el jugador uno, y *b* el pago para quien mueve segundo, designado II. Si I elige C, entonces a II le toca elegir entre las opciones de explotar (E), donde I recibe un pago de –1 y II recibe un pago de 3, y cooperar (R) donde ambos reciben pagos de 2. Todos estos pagos son conocimiento común –esto es, son conocidos por ambos jugadores, y se saben conocidos-**.**

En el juego de confianza tal como fue representado, hay un equilibrio para los jugadores racionales –a saber, N-. En efecto, I sabe que II, si es racional, elegirá E por sobre R, y recibirá así un pago de 3 en lugar de 2. Pero en E, I recibe –1, mientras que recibiría cero en N: en consecuencia, I no confiará. Y esto revela el problema. Como el dilema del prisionero, este juego tiene un equilibrio que es ¨Pareto-dominado¨ por otro resultado técnicamente posible, R. Esto es, el equilibrio en N implica pagos menores para ambos jugadores de los que cada uno obtendría en R, pero R se vuelve inaccesible en virtud de la racionalidad de II.

[Tal vez valga la pena señalar aquí que al atribuir la inaccesibilidad de R a la racionalidad de II, me estoy comprometiendo con un enfoque particular de la racionalidad (el enfoque habitual del economista) que es filosóficamente discutido. Vale la pena notar los temas que están en juego aquí. Si es verdad que la racionalidad en el campo de la acción hace que la vida de uno vaya menos bien de lo que iría si uno fuera,

de modo identificable, ¨irracionalmente¨ confiable, entonces uno tiene razones racionales para volverse (irracionalmente) confiable. Es claro que nuestra terminología es inadecuada aquí. Necesitamos distinguir la racionalidad A (racionalidad en el campo de la acción) de la racionalidad D (racionalidad en el campo de las disposiciones, donde una disposición debe ser entendida como un modo de calcular que podría sopesar argumentos distintos de los pagos objetivos al decidir la conducta). Lo que podríamos decir, entonces, es que es racional-D no elegir la disposición de racionalidad-A. Es, creo, inútil argumentar (como hace Gauthier) que si el jugador II se forma la *intención* de elegir C en primer término, esta formación de intención hace que R sea la elección racional de acción en el segundo período. Me parece que es, en el mejor de los casos, engañoso decir que R es la opción ¨verdaderamente racional¨ para II *porque la intención de hacer R fue formada racionalmente*. Quiero insistir en que la *acción* racional para II es E (me pongo, tal como lo entiendo, del mismo lado que Derek Parfitt en esta cuestión). En el mismo sentido, querría rechazar las sugerencias de Elster (1981; 1983, por ejemplo) de que el problema surge aquí debido a la racionalidad *imperfecta*. Esta descripción sugiere que si II fuera *completamente* racional, II elegiría R y no E. Desde mi punto de vista, es un rasgo característico de la racionalidad que la elección racional es prospectiva y autointeresada**:** sólo si E *no* fuera la mejor elección desde la posición de quien elige segundo cuando él está realmente en el nodo de elección número 2 podríamos decir que R era la elección racional.]

Fig.1: El juego básico de la confianza

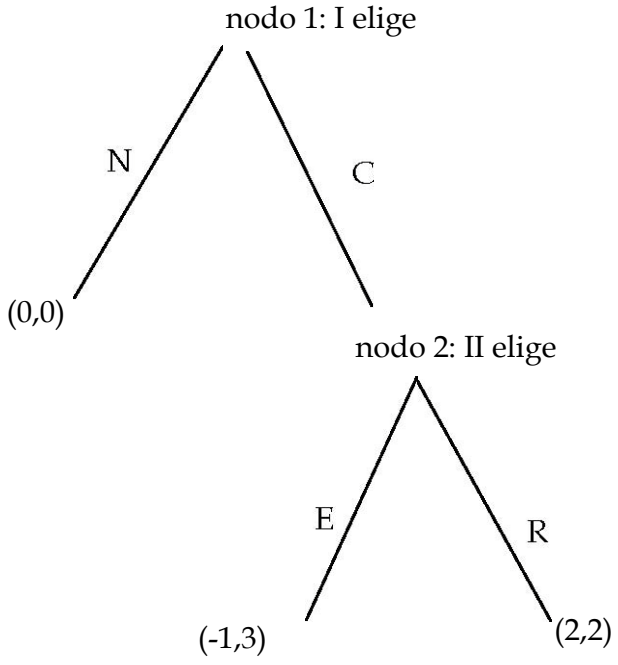
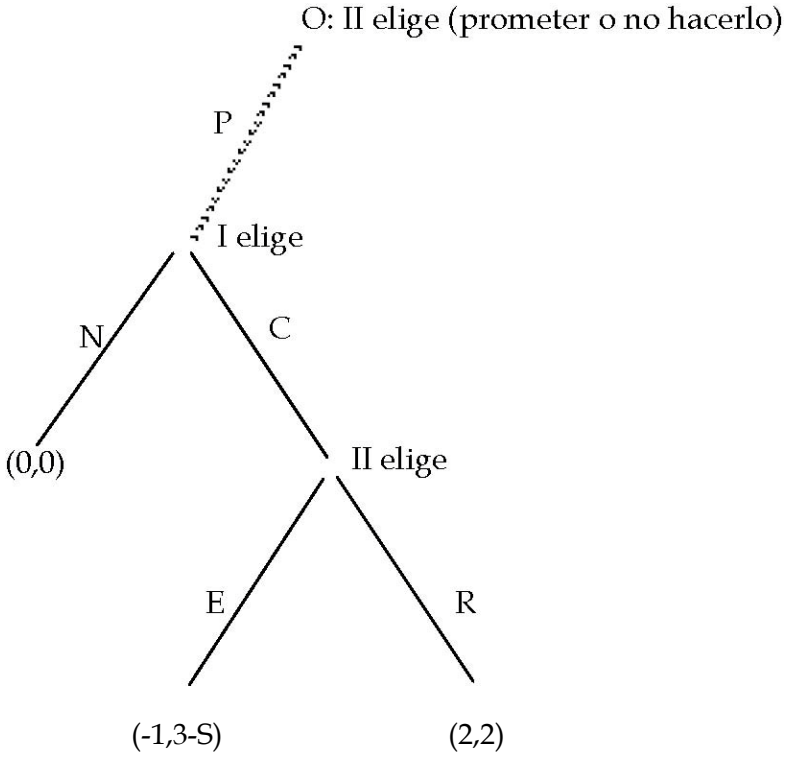


Fig.2: El juego básico de la confianza <u>cuando el segundo jugador es confiable</u>



Por supuesto, si la Fig. 1 no fuera una representación completa de la interacción entre I y II –si, por ejemplo, la interacción está en curso y el hecho de que II elija E en la jugada en curso induce a I a elegir N en el futuro, entonces podría haber una razón para que II elija R -. Pero este punto, familiar para los economistas como ¨la disciplina del comercio continuo¨, aunque es enteramente válido y útil en sí mismo, no viene *al caso* en el juego de la confianza. Aun si el juego de la confianza es jugado nuevamente, y las reputaciones de los jugadores están públicamente disponibles, siempre habrá algunas interacciones que tengan la estructura del juego tal como está representado: siempre habrá *algunos* intercambios, en potencia mutuamente beneficiosos, que no serán consumados por agentes completamente racionales (porque la acción racional para quien mueve segundo le dictará que elija E).

La noción de racionalidad-D sugiere, sin duda**,** una solución para el predicamento de la confianza – una ¨solución¨ que tal vez nos acerque a nuestras intuiciones sobre la confianza. Puede, hemos dicho, ser racional para II al nivel de la elección disposicional que II elija una disposición a mantener las promesas. Para una persona así, lo pasado no está pisado: una promesa hecha, un compromiso contraído, serán, podríamos decir, costosos de romper. Caracterizaremos a esa persona confiable como una persona cuyos pagos son aquellos mostrados en la figura 2 donde *S* es un ¨parámetro de conciencia¨ y toma un valor mayor a

1. Si queremos mantener el espíritu de la prospectividad instrumental, característico de la racionalidad, podríamos pensar en *S* como la carga de vergüenza y culpa que II soportará si II no cumple con un compromiso contraído (deberíamos poner énfasis sin embargo en que *S* no tiene que ser interpretado de ese modo instrumental). (De hecho, la Fig. 2 es una versión truncada del juego real: para que sea completo, necesitaríamos mostrar una elección anterior (período 0) hecha por II en que prometa o se comprometa (estrategia P))**.**

Sin embargo, es un rasgo crucial de la racionalidad-D que a II sólo le convendrá elegir la disposición confiable si el hecho de que lo haga aumenta la probabilidad de que I confíe en él. Si I confía ¨a ciegas¨ -con independencia de si tiene o no razones para creer que II es confiable o no- entonces a aquellos de los que mueven en segundo lugar que no sean confiables les irá mejor que a los que sean confiables: entonces no será racional-D ser confiable.

Consecuentemente, un factor crucial en la historia de la confianza racional(-D) es el grado de ¨translucidez¨ en las relaciones entre las personas –el grado en el que quienes mueven primero pueden identificar las disposiciones de quienes mueven segundo-. Para tomar el caso límite, si quienes mueven primero pueden identificar perfectamente a aquellos de los que mueven segundos que son confiables, entonces a todos los agentes les convendrá adoptar la disposición a ser confiable.

O, al menos, este será el caso si las disposiciones *pueden* ser racionalmente elegidas de este modo. La idea de racionalidad-D simplemente presupone que la elección de disposiciones está disponible libremente –que *podemos* transformarnos por asimilación cultural de manera confiable para mantener nuestros compromisos- pero bien podría haber dificultades aquí. Supóngase, como la racionalidad-D parecería requerir, que la razón por la que tratamos de adaptarnos por asimilación cultural a la disposición de mantener nuestros compromisos es que creemos que así seremos recompensados al ser sujetos de confianza. Ahora, sin embargo, precisamente esa misma consideración nos dará razones para *no* mantener nuestros compromisos cuando nos toque elegir si vamos a *actuar* de modo confiable. Si debemos genuinamente sentir vergüenza y culpa por romper nuestras promesas, debe ser porque creemos que romper promesas está mal –no porque creemos que el pensar que romper promesas está mal será bueno para nosotros-. Un ejemplo de John Broome es revelador en este contexto. Suponga que usted está enfermo. Y suponga que su velocidad de recuperación depende hasta cierto punto de su estado mental. Específicamente, suponga que, si usted cree que se recuperará en n días, se recuperará en (n+2) días

donde n 1. Claramente, es mejor que usted crea que se curará mañana, porque entonces se curará en tres días. Pero el deseo de curarse tan pronto como sea posible no puede proporcionar una *razón* para creer que usted se curará mañana: la única razón correcta para que usted crea que se curará mañana es que dé la casualidad de que usted crea que la proposición relevante (que usted se curará mañana) es *verdadera*. Bien podría ser que la confiabilidad sea así. Tal vez el único modo de ser confiable es tener una creencia genuina de que romper promesas es moralmente incorrecto. Y las creencias de esta clase no pueden simplemente ser manufacturadas de manera autointeresada. O eso podría afirmarse. Si la disposición a ser confiable depende de ciertas creencias morales, y si tales creencias son inaccesibles para la elección racional son temas que necesitamos notar, pero que no podemos tener esperanzas de responder aquí. Por mi parte, me inclino a pensar que probablemente *hay* un componente moral irreducible implantado en la ¨elección¨ disposicional plausible. Sin embargo, aun si esto no fuese así, tener las creencias morales relevantes ciertamente parecería hacer su confiabilidad más robusta.

Si la confiabilidad es racionalmente-D inaccesible de esta manera, podríamos sin embargo suministrar una explicación del surgimiento de la confiabilidad en términos evolucionistas. Las explicaciones evolucionistas no necesitan apelar a la intencionalidad del agente. Ciertamente esperaríamos que a los tipos confiables les vaya mejor que a los que son meramente racionales-(A), si suponemos que el hecho de que les vaya mejor en términos de pagos se correlaciona bien con las características de supervivencia. Aun aquí, por supuesto, sigue siendo necesario que las disposiciones de los agentes sean ventajosa: quienes mueven primero deben poder diferenciar con cierto grado de precisión a los tipos confiables de los que no son confiables. De otro modo, ser confiable no aumentará los pagos. ¿Pero qué tal si el grado necesario de translucidez está ausente? ¿Qué tal si, del mismo modo en que nuestros poderes para elegir a los buenos muchachos como jueces es inadecuado, también nuestra capacidad de elegir buenos muchachos como socios comerciales es imperfecta? ¿Qué**,** si es que podemos

hacer algo, podemos hacer?

3. **Entran en escena los tribunales, con un elogio:**

Tal vez sea natural que cualquiera que llegue al predicamento de la confianza –al menos cualquiera que no esté profundamente transformado por asimilación cultural a la lógica de la teoría de la elección pública- verá a ese predicamento como la razón central para el derecho contractual. Precisamente porque los agentes no pueden contraer compromisos creíbles en montones de casos donde sería provechoso para ellos hacerlo, necesitamos el derecho contractual, apoyado por poderes de ejecución, para hacer esos compromisos creíbles. Sin embargo, si aplicamos la lógica de la teoría de la elección pública –y el principio de simetría motivacional en particular- entonces inmediatamente nos enfrentamos con el problema del q*uis custodiet*. Si los jueces y la policía deben ser considerados completamente confiables, entonces también deben serlo los comerciantes comunes. Y si los comerciantes comunes son completamente confiables, no necesitamos los tribunales. Alternativamente, si los comerciantes son habitualmente personas que *no* son confiables, entonces meter por la fuerza a los tribunales en el predicamento de la confianza es simplemente admitir a otro jugador, con poderes especiales, que es de pies a cabeza tan racionalmente autointeresado como todos los otros. No parece haber ninguna razón *a priori* por la que este tercer jugador habría de ejecutar de manera confiable los contratos entre los otros dos. Seguramente simplemente expropiará todas las rentas que estén disponibles dentro de los límites (si es que hay alguno) de sus poderes asignados. En estos dos casos, en cualquiera de los extremos de un hipotético espectro motivacional, los tribunales no parecen agregar nada. No pueden hacer una tarea útil si todos son confiables (porque no los necesitamos) o si todos son personas en las que *no* hay que confiar (porque los jueces no ejecutarán la ley de manera confiable). Pero ese no es el fin de la historia. En al menos

algunos casos intermedios –casos en los que hay tipos confiables y otros que no lo son - los

genuinos de ejecución, pueden de hecho hacer una tarea positiva. Pronto especificaré exactamente cuál es esa tarea positiva.

Para aclarar la clase de suposiciones acerca del conocimiento de otros jugadores que tengo en mente aquí, considérese un contexto evolucionista en el que inicialmente hay una proporción P de tipos confiables, donde $0<P<1$. Los jugadores son escogidos al azar de la población para jugar el juego básico, y sus roles (como quien mueve primero o segundo) también son determinados al azar. Se supone que P es conocido por todos los jugadores, así que cada uno de los que mueve primero sabe que quien mueve segundo será confiable con una probabilidad P. Para asegurar que el principio de simetría motivacional se cumpla, los jueces también se eligen al azar del mismo grupo que los jugadores; de acuerdo con esto, la probabilidad de que un juez sea confiable es la misma probabilidad P. Los jugadores juegan el juego de la confianza una vez, y entonces son reubicados para un nuevo enfrentamiento al azar en juegos subsiguientes. Los jugadores son anónimos, así que no hay posibilidad de construir una reputación.

Aquí haré una suposición crítica acerca del papel de los jueces. Esta es que el papel del juez es *reactivo* en el sentido de que el juez sólo puede ser invitado a jugar en el juego sustantivo por uno de los jugadores. Entonces, mientras los jueces tienen un poder genuino para que sus juicios sean ejecutados y aunque este poder los habilitará para recaudar rentas de los jugadores, los jueces sólo llegan a ejercer este poder por invitación.

¿Cómo actuarán los jueces? Supondré que los jueces confiables decidirán en favor de la parte explotada, si es que hay una, y ejecutarán una redistribución del explotador al explotado para asegurar que el explotado reciba lo que le fue prometido. Entonces, en nuestro juego de la confianza original, por ejemplo, si quien mueve segundo escoge E y el segundo a quien mueve primero, para asegurar que quien mueve primero reciba el pago de 2 que le fue prometido; esto dejará a quien mueve segundo con un pago de cero. Supongo, además, que los jueces que no sean confiables decidirán los casos al azar; pero

sistemáticamente se apropiarán de rentas de los jugadores, un monto X del jugador I, y un monto Y del jugador II**.** Entonces si el resultado es E y el proceso de resolución de controversias es activado, en el caso en que el juez *no* es confiable, la estructura de pagos será:

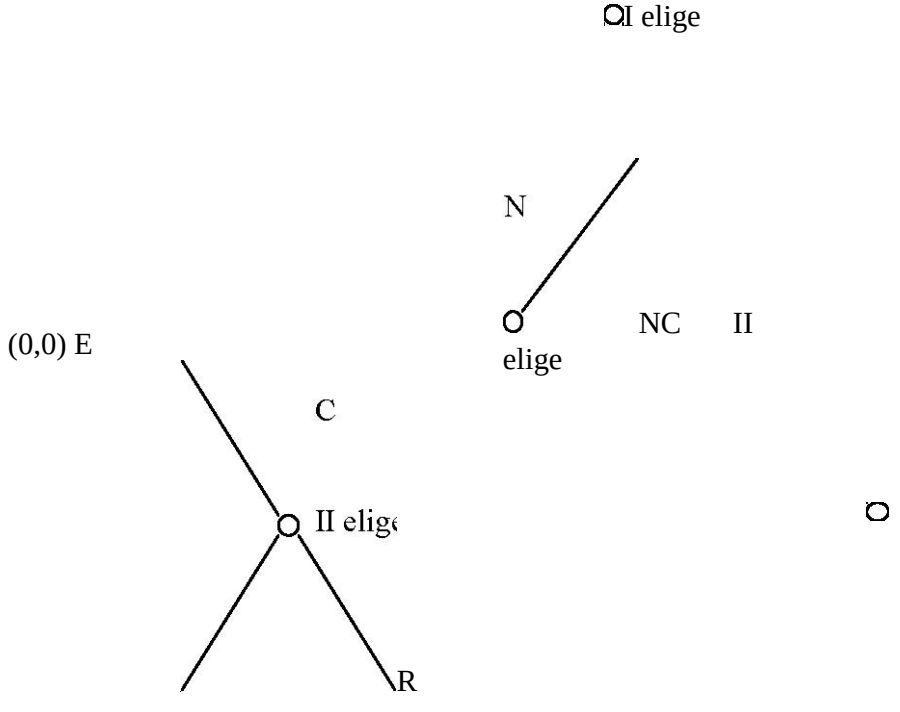
½ (2-1)-*X* = [½ -*X*] para I y ½ (3+0)-*Y* [=3/2 –*Y*] para II

Ahora bien, hay una simplificación importante a mano aquí. Dado que siempre hay una posibilidad (1-P) de tener un juez que no sea confiable, hay un costo esperado para los jugadores que invoquen el proceso de resolución de controversias. Nótese que quien mueve primero y es explotado es el único jugador que puede alguna vez ganar algo del proceso de resolución de controversias. Por lo tanto, el único caso en el que el proceso de resolución de controversias será activado será cuando E es el resultado y la invitación proviene de quien mueve primero y es explotado. Por supuesto, si X e Y son grandes en relación a los pagos (y no hay nada aún que indique por qué deberían ser ilimitados), incluso quien mueve primero y es explotado podría no activar el proceso de resolución de controversias. Pero la simplificación aquí es que, para determinar si el proceso de resolución de controversias será invocado o no, sólo necesitamos saber el cálculo de quien mueve primero y es explotado.

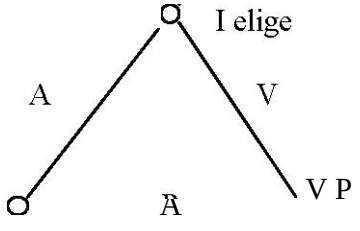
como una versión expandida del juego básico. Efectivamente, lo que sucede es que si E emergiese como el resultado del juego, el primer jugador, I, tiene otra elección: él puede o bien prestar su aquiescencia, A, y aceptar el resultado (-1, 3); o bien activar el proceso de resolución de controversias, V. En el último caso, el resultado es (2,0) con probabilidad P (*i.e.*, cuando uno tiene un juez confiable) y (1/2 –*X*, 3/2 – *Y*) con una probabilidad (1-P) (*i.e.*, cuando el juez no es confiable).

Todo el árbol del juego está representado en la Fig. 3 (aunque mostramos la opción de invocar el proceso de resolución de controversias sólo para el caso en que quien mueve segundo no es confiable y elige E). La contribución del proceso de resolución de controversias puede entonces ser evaluada a través de una comparación entre el juego *con* el proceso y *sin* él.

Fig. 3



(1-p) (-1,3) (2,0) (1/2 – X, 3/2 – Y)



El jugador I razonará que el jugador II resultará confiable en P casos y no será confiable en (1-P) casos. El

retorno esperado de elegir C es entonces:
(C) = 2.P + (-1) (1-P) = 3P – 1 Entonces, I confiará si P > 1/3 y no confiará si P < 1/3 Los pagos del juego son, entonces: para P > 1/3 (3P-1, 3-P), total 2+2P para P < 1/3 (0, 0), total cero Nótese, además, el efecto sobre el valor del ¨equilibrio evolutivo¨ de P. Si el valor inicial de P, denotado por P1, es mayor que 1/3, entonces I confiará racionalmente. Por consiguiente, a los tipos que no sean confiables y estén en el papel de quien juega segundo les irá mejor que a los tipos confiables que estén en el papel de quien juega segundo; entonces P disminuirá, hasta que alcanzamos un punto donde P<1/3, donde nadie confiará. En este punto, a los tipos confiables y a los que no lo son les va igualmente mal y P no seguirá disminuyendo. Por sobre este punto, el hecho de que haya tipos confiables no hace bien alguno: simplemente no hay suficientes tipos así para inducir a quienes mueven primero y son racionales a confiar, y el pago para todos los jugadores en el juego es cero. El caso en presencia de los tribunales es algo más complicado. El tema central aquí es si, y bajo qué circunstancias, I elegirá V en lugar de A. Dado que si es racional para I elegir A, II conocerá ese hecho, y la solución para el juego extendido será idéntica a aquella en el caso del juego sin resolución de controversias. Por consiguiente, nos enfocamos sobre la elección de quien mueve primero de si recurrir al proceso de resolución de controversias

(V) o no hacerlo (A). Quien mueve primero elegirá V sólo si

i.e., $P \square (2X - 3)/(2X + 3)$. Significa que $Pv = (2X - 3)/(2X + 3)$

Este parámetro, Pv, representa el valor de umbral de P, tal que si P>Pv, entonces I irá a juicio. Para propósitos ilustrativos, podríamos notar que Pv es bastante menor que 1 aún para valores de X bastante grandes en relación a los pagos en el juego básico: para X=8.5, P=.7; para X=10, P=.75; para X=3.5, P=.4. Obviamente, a medida que X se vuelve extremadamente grande, entonces Pv se acerca más y más a la unidad, tal como esperaríamos. Si las rentas que un juez que no es confiable puede extraer son muy grandes, la proporción de personas confiables requerida para que I racionalmente recurra al procedimiento de resolución de controversias tendrá que ser proporcionalmente más grande.

En este punto, nótese lo que sucede si esta condición es satisfecha. El jugador II *sabrá* que este es el caso. Y el jugador II puede actuar para evitar el resultado V eligiendo R en el período anterior. ¿Será racional que lo haga? Sí**,** si 0.p + (1-p) (1½ -Y) <2. Y esta condición siempre será satisfecha bajo valores plausibles de Y. (De hecho, en este ejemplo, sólo requerimos que Y sea mayor que ½) En este sentido, el ámbito para que un juez que no es confiable obtenga rentas significativas es una ventaja, ya que induce a los jugadores que no son confiables a cumplir con sus compromisos para así evitar tener que enfrentar a los tribunales. Si la proporción de jueces confiables es suficientemente alta para inducir a quien juega primero y es explotado a invocar el proceso de resolución de controversias, entonces quienes muevan segundos y no sean confiables racionalmente actuarán para hacer que tal invocación sea innecesaria *actuando* de manera confiable.

controversias donde P>Pv. Aquí, todo el mundo –los que son confiables y los que no lo son- cumplirán con sus compromisos. Los que muevan al final sabrán esto (porque saben P) y por lo tanto confiarán. Entonces el resultado es:

Para P>Pv: (2; 2) total 4 Para Pv>P>1/3: (3p-1; 3-

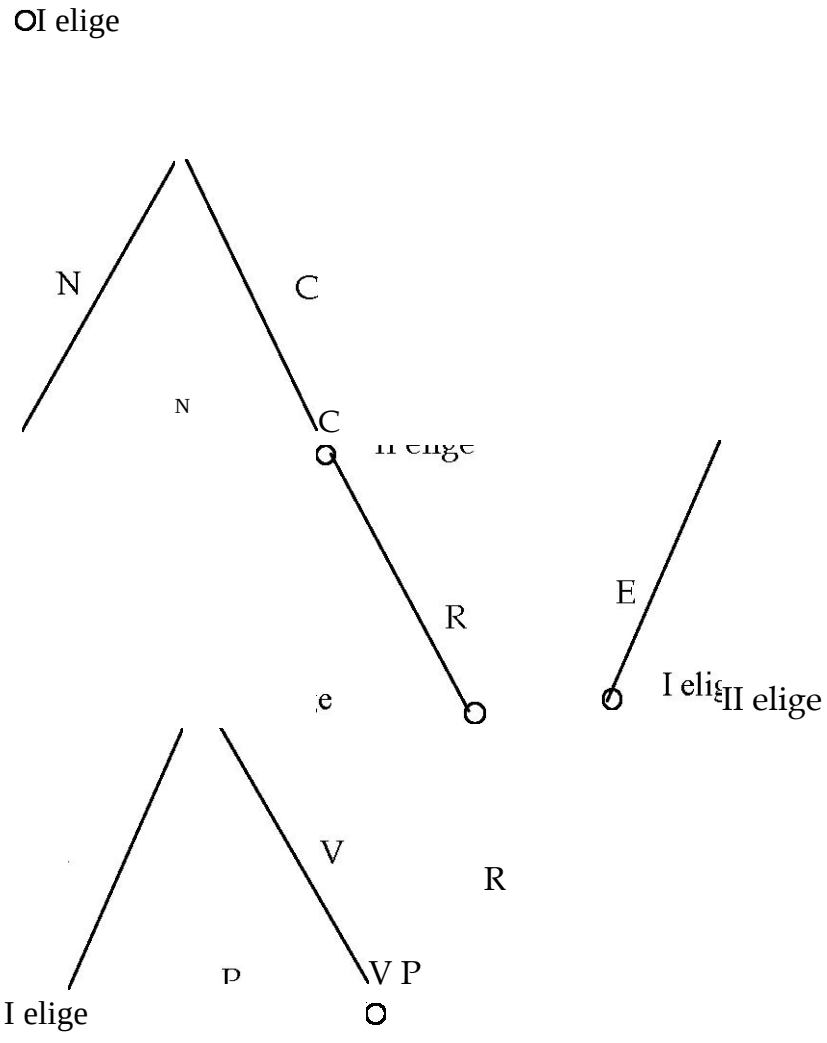
p) total 2+2p Para P< 1/3: (0; 0) total 0

Además, para los casos donde el valor inicial de P excede Pv, los tipos confiables y los que no son confiables son equivalentes en cuanto a sus conductas, entonces en particular los tipos que no son confiables no tienen ninguna ventaja evolutiva. Por lo tanto no hay ninguna razón para pensar que los tipos confiables serán excluidos en el equilibrio evolutivo. Por supuesto, a los tipos confiables les va bien, aunque mucho menos que a los tipos que no son confiables en sus papeles como *aquellos encargados de resolver las controversias*; pero el carácter peculiar de esta forma de estructurar el modelo de resolución de controversias es que el proceso de resolución nunca es invocado realmente. Dado que el proceso de resolución de controversias impone costos netos asimétricos sobre los jugadores que no son confiables, ellos encuentran conveniente mantener sus compromisos, precisamente para evitar enfrentar la resolución de controversias. Los jueces nunca llegan a actuar, por lo tanto la diferencia de pagos para los jueces que no son confiables es irrelevante.

Obviamente, una pregunta crítica en toda esta historia es cuán rigurosa es la valla que Pv representa: si Pv está razonablemente cerca de la unidad, entonces el proceso de extremas donde P es alto. Este valor crítico, Pv, depende crucialmente de X, las rentas que un juez que no es confiable puede extraer de quienes mueven primero. ¿Qué valor de X es plausible? En particular, ¿cómo puede determinarse X?

Nuestra propia construcción sugiere un mecanismo por el cual X *puede* ser determinada. Supóngase que imponemos otra capa de resolución de controversias en el proceso desplegado hasta ahora: esto es, permitimos que los jugadores sustantivos apelen la decisión del tribunal de primer orden. En este nivel de apelación, sin embargo, no hay una restricción posterior, entonces podríamos esperar que los jueces que no son confiables en este nivel se apropien de rentas verdaderamente máximas. Supóngase que lo máximo que puede extraerse de los jugadores es M. Podemos representar la situación en la forma del árbol de juego en la Fig. 4**.** Este es simplemente Fig.3 con la opción extra para I (quien es el más relevante aquí) de apelar la decisión si es que a él le toca un juez que no es confiable en el proceso de resolución de la primera ronda. Las opciones son o bien apelar [representado por L] o aceptar el veredicto [representado por G]. Nótese que el juez de primera instancia que no es confiable esperará perder en caso de una apelación, sin importar que el juez de segunda ronda sea o no confiable. Si el juez de la segunda ronda es confiable, corregirá la decisión de la primera ronda, reintegrará a I el monto de 2 prometido y reintegrará a sus legítimos propietarios las rentas de las que el juez que no fue confiable se apropió indebidamente. Si el juez de segunda instancia no es confiable, se apropiará de *todas* las rentas disponibles, no sólo de I y II, *sino también del juez de la primera ronda*. (Los terceros términos en los pagos en los pagos finales del período 4 indican los pagos a los jueces de primera instancia). El punto crítico aquí es que no será racional para ese juez apelación, ese juez perderá.

Fig.4:



(0,0)
E

(2,2)

A

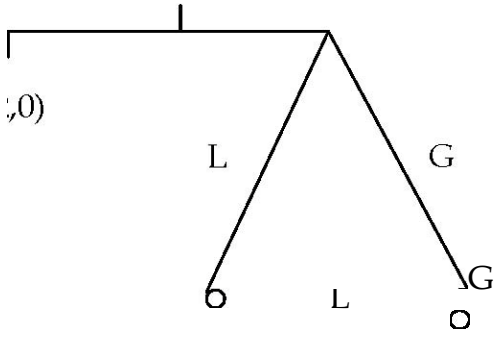
1-P I elige
(-1,3)

(2,0)

(1/2 – X, 3/2 – Y) (2,0) (1/2 – M, 3/2 – M)

[M>x,y]

Por consiguiente, deben darse la condiciones de que: y 0.p + (1-p) (3/2 – M) < 1 ½ - Y para que II no apele Entonces la condición sobre X se vuelve: p< (M – X + ½) / (M+2) que es aproximadamente X/M < 1-p Y la condición sobre Y se vuelve Y/M < 1-p+ (3/2M)



Nótese que la condición es más rigurosa sobre X que sobre Y, entonces hay razones para pensar que los jueces que no son confiables, sujetos a una posible apelación, extraerán rentas mayores de explotar a quienes mueven segundo que a quienes mueven primero y son explotados. Más pertinente, tal vez, es que las rentas tomadas por un juez de primera ronda que no es confiable de manera racional serán *limitadas*, tal vez a un tercio del total de las rentas apropiables, para asegurar así que apelar no sea racional para los jugadores.

4. Conclusión:

Esta porción del argumento completa la idea. La organización institucional que se ofrece aquí consiste de jueces escogidos al azar de la misma población que los jugadores, con un mecanismo de apelación (que implica a otro juez escogido al azar) disponible para los jugadores. El juicio del juez es ejecutable: los jueces tienen poder genuino hasta ese punto. Pero el proceso de resolución de controversias no puede ser activado por nadie más que los propios jugadores. El proceso es reactivo en ese sentido**.**

los contratos, para valores razonables de la proporción de personas confiables –valores, ciertamente, considerablemente menores al 100 por ciento-. Además, el proceso sirve para prevenir la erosión de la confiabilidad en circunstancias donde la confiabilidad de otro modo *se erosionaría*. Estos son logros mayores. Es verdad que asegurar esos logros depende de que la proporción de personas confiables en la sociedad no sea demasiado baja –pero no requiere que los hombres sean ángeles-. Bien lejos de la rectitud universal, para repetir la cita anterior de Hamilton, puede haber una porción de honor entre la humanidad que puede ser la base de nuestra esperanza.

Para aquellos a los que no les gusta la aritmética, podría ser útil afirmar lo que está en juego en el argumento precedente, despojado de todo el detalle analítico. Expuesta de manera simple, la historia es esta. El problema de la confianza surge porque las partes que actúan racionalmente no pueden contraer compromisos creíbles –no hay medios disponibles a través de los cuales una persona que elige en todo momento aquella acción que hace que su vida vaya mejor para ella pueda cumplir con las promesas hechas-. Lo que los tribunales, bajo la descripción dada aquí, ofrecen es un mecanismo efectivo para que quienes mueven primero hagan amenazas creíbles –al menos bajo ciertas circunstancias-. La razón por la que los tribunales ofrecen esta facilidad es que quienes mueven primero y son explotados – quienes confían en alguien que no cumple con promesas hechas- pueden apelar a los tribunales; y la amenaza de esa apelación inducirá a quienes mueven en segundo lugar y son racionales a cumplir. El resultado se seguirá aun si los tribunales son algo corruptos, a condición de que no sean *totalmente* corruptos –esto es, a condición de que no *todo* hombre sea un bribón-. Sin embargo, requiere que los tribunales sean *reactivos* –que sean activados que los agentes sean capaces de distinguir a los tipos confiables de los que no son confiables, en la selección de socios comerciales o en la selección de personas para el papel de jueces. En ese sentido, no implica ninguna violación del principio de simetría motivacional: los jueces no son diferentes (y específicamente no son mejores) que los jugadores comunes. En este sentido también, la

pregunta sobre quién vigilará a quienes nos vigilan es vista como una pregunta de segundo orden –y ciertamente no es necesario que sea un desafío demoledor a las instituciones de poder delegado-. En el modelo presentado, los tribunales pueden hacer una tarea normativamente relevante aun cuando los vigiladores no son vigilados. La creación de su presencia, en sí misma, puede ser suficiente.

Existen dos comentarios particulares sobre detalles del modelo –dos posibles críticasde las que soy consciente y para las que no tengo ninguna respuesta satisfactoria, al menos hasta este momento. Uno se relaciona con la naturaleza de la confiabilidad que tomé como un presupuesto. He tratado aquí a la confiabilidad como una disposición. Esto es, *S* es suficientemente grande para asegurar el comportamiento confiable de un tipo confiable no obstante la tentación. Esto significa que la proporción de jueces confiables es idéntica a la proporción de jugadores confiables, a pesar del hecho de que los agentes en el papel de jueces enfrentan tentaciones más sustanciales de explotar que las que aquellos mismos agentes enfrentan en el papel de jugadores comunes. Concedo que esta suposición es inverosímil. Me inclino a creer que hay, como podríamos exponerlo, una curva de demanda de pendiente descendente para la moralidad como para otras cosas. El modo en que he modelado las disposiciones es, en este sentido, excesivamente rígido. Sería posible en análisis mucho más complicado y no alteraría, creo, los resultados cualitativos.

La segunda crítica se relaciona con la suposición de tribunales reactivos. ¿Quién, uno podría preguntar, debe ejecutar ese requisito? ¿No nos regresa este requisito al problema del *quis custodiet* una vez más? Posiblemente. Mi respuesta aquí es simplemente observar que algunas organizaciones institucionales requieren una mayor virtud humana que otras. Los economistas tienen una inclinación profesionalmente fundada a pensar que *cualquier* requisito de virtud humana es una señal automática de falta de viabilidad: este es el mensaje de la observación de Hume citada al comienzo. Pero aquí me pongo explícitamente del mismo lado que Hamilton. Tal como he observado en otro lado, uno no tiene que creer que algo es *infinitamente* escaso para creer que economizarlo vale la pena. Tal como este modelo ilustra, si uno adopta una línea menos rigurosa (y en mi visión más plausible) sobre la virtud humana –una línea más cercana, digamos, a Adam Smith que a Bernard Mandeville- entonces se ofrecen posibilidades para el diseño institucional que en el cuadro motivacional más extremo serán simplemente declaradas ¨improcedentes¨, por decirlo de algún modo**.**

Finalmente, y ya que estamos en la cuestión de problemas ¨fuera de los tribunales¨, es un rasgo interesante del cuadro que he esbozado que nunca se recurre a los tribunales en realidad. Los tribunales hacen su trabajo creando amenazas creíbles, más que actuando. Y me parece que este resultado probablemente sea robusto en cualquier modelo en el que la acción racional es supuesta. Esto es, en modelos de actividad de los tribunales con jugadores esencialmente racionales, es probable que sea un rasgo que nunca se recurra a los tribunales, excepto por accidente –y esto bajo una variedad de detalles institucionales

mucho más amplia que los esbozados aquí-. Si el papel principal de los tribunales es alterar los incentivos para asegurar resultados particulares, entonces la *racionalidad* debería asegurar que1os procesos judiciales no sean invocados. En este sentido, el hecho de que los tribunales sean ¨reactivos¨ en los términos usados aquí es un rasgo extremadamente importante de todo el argumento. Una defensa general de reglas apropiadas de ¨legitimación¨ se sigue más o menos directamente del argumento previo.

BIBLIOGRAFÍA

Brennan, Geoffrey & James Buchanan (1985) *The Reason of Rules*, Cambridge University Press, Cambridge. Brennan, Geoffrey & Alan Hamlin (2000) *Democratic Devices and Desires* Cambridge University Press, Cambridge.

Brennan, Geoffrey, Werner Guth and Hartmut Kliemt (1998) '*Trust in the Shadow of the Courts*' [mimeo] Hume, David (1985) *Essays Moral, Political and Literary* (ed E. Miller) Liberty Press, Indianopolis.

Madison, James, Alexander Hamilton and John Jay (1788/1987) *The Federalist Papers* (ed. I. Kramnick), Penguin Books, Middlesex. Mueller, Dennis (1989*) Public Choice II*, Cambridge University Press, Cambridge. Sally, David (1995) 'Conversation and Co-operation in Social Dilemmas' *Rationality and Society*, p58-92.