Justicia de Mano Propia: Un Experimento de Castigos de Terceros[1]

Freddy H. Mendoza (freddy.mendoza@correounivalle.edu.co)[2]

Lina Restrepo-Plaza (lina.restrepo@correounivalle.edu.co)[3]

Resumen: Las continuas fallas en las instituciones formales para la administración de la justicia han promovido el desarrollo de instituciones informales basadas en la generación de castigos en manos privadas. Esta investigación propende por evaluar el rol de la solidaridad, la indignación y la venganza en el proceso de toma de decisiones de sancionar. Con tal propósito, hemos realizado un experimento basado en el *trust game* con dos adaptaciones: incorporando el *cheap talk* y la posibilidad de castigo. En nuestro diseño experimental manipulamos, cuidadosamente, la presencia de castigo, quién castiga y la estructura de pagos de los jugadores. Encontramos que la probabilidad de castigo y las decisiones de inversión no se ven modificadas por nuestros tratamientos, sin embargo, sí modifica las actitudes recíprocas. Concretamente, identificamos que la posibilidad de que un agente ejecute una sanción motivada por la venganza inhibe las actitudes egoístas de quienes pueden apropiarse de las ganancias de la inversión, pero la indignación y la solidaridad no parecen surtir ningún efecto.

Palabras Clave
Reciprocidad, Confianza, Castigo de terceros, Experimento de laboratorio, Juego de confianza

Clasificación JEL C91 G11 F51

[1] Este documento de trabajo fue financiado, parcialmente, por el programa ``Inclusión productiva y social: programas y políticas para la promoción de una economía formal", código 60185, que conforma Colombia Científica-Alianza EFI, bajo el Contrato de Recuperación Contingente No. FP44842-220-2018.

[2] Economista, miembro del grupo de investigación E-Socials: Experimental Social Sciences and Behavioral Change de la Universidad del Valle. Profesional de Planeación. Universidad Libre, Cali, Colombia
[3] Profesora asociada del departamento de Economía de la Universidad del Valle. Directora del grupo de investigación E-Socials: Experimental Social Sciences and Behavioral Change. Perteneciente a la Red de Investigadores del Banco de la República.

The Law into One´s Own Hands: a Third-Party Punishment Experiment[4]

Freddy H. Mendoza (freddy.mendoza@correounivalle.edu.co)[5]

Lina Restrepo-Plaza (lina.restrepo@correounivalle.edu.co)[6]

Abstract: Formal institutions in charge of justice administration often fail, which in turn promotes the generation of private sanctioning actions. This paper evaluates the role of three mechanisms that might explain the law into one's hand: solidarity, vengeance, and resentment. We ran an adaptation of the trust game that entails a cheap talk and punishment possibilities. We carefully manipulate when and who is punishing, and the players' payoff structure. We find that our treatments do not affect punishment and investment decisions, but they have a reciprocity effect. More concretely, we identify that vengeance-based punishment is more effective in promoting reciprocity than other motivations.

Key Words:
Reciprocity, Trust, Third-party, Punishment, Lab-experiment, trust game.

JEL Classification: C91 G11 F51

---

[4] The Borradores de Economia series is a publication of the Economics Studies Unit at Banco de la República. The documents here published are provisional, and the authors are fully accountable for the opinions and eventual mistakes. The content of this document compromises neither Banco de la República nor its Board of Directors.

*This working paper was partially funded by the Colombia Científica-Alianza EFI Research Program, with code 60185 and contract number FP44842-220-2018, funded by The World Bank through the call Scientific Ecosystems, managed by the Colombian Ministry of Science, Technology and Innovation.

[5] Economist member of the research group E-SocialS: Experimental Social Sciences and Behavioral Change at Universidad del Valle. Practitioner at Universidad Libre, Cali, Colombia.

[6] Associate professor of the Department of Economics at Universidad del Valle. Director of the research group E-SocialS: Experimental Social Sciences and Behavioral Change. Member of Banco-de-la-República´s Research Network.

1. Motivación

La confianza se configura como un conjunto de actitudes, deseos, creencias, emociones y expectativas que constituyen con frecuencia la esencia del ser humano. Transgredir la confianza de un individuo tiene resultados negativos que afectan no solo al directamente involucrado, sino también al entorno social. Por esta implicación, la construcción, preservación o vulneración de la confianza y su implícita reciprocidad se convierten en unos de los principales focos de estudio de las disciplinas que analizan el comportamiento humano. La economía es una ciencia social que estudia las decisiones de asignación de recursos y las interacciones en el mercado. La neoclásica lo hace desde el paradigma de la escasez y la racionalidad.

¿Cómo podemos, entonces, reconciliar el egoísmo que se deriva de la explicación de la Escuela neoclásica de economía con situaciones recurrentes que exigen la confianza y cooperación entre agentes? La confianza y la reciprocidad constituyen un concepto extraño dentro de la lógica económica, ya que las interpretaciones, hasta el momento, se han basado en un visión puramente individualista y no colectiva. Sin embargo, a finales del siglo pasado se han realizado estudios que contradicen la teoría neoclásica sobre el comportamiento humano, como el experimento de Berg et. al (1995), en el que por medio de un juego de inversión (Investment Game) se observa que la cooperación entre dos individuos denominados *investor or depositors*, se encuentra entre el 40% y el 60%. Con base en el experimento, se puede deducir que en la vida real hay una gran variedad de situaciones en las que los individuos parecen tener inclinaciones colectivistas y en los que la confianza es la decisión óptima.

Algunas de la razones para tener dichas inclinaciones pueden ser que los individuos creen que hacer parte de un colectivo social implica responsabilidad con el otro y colaboración mutua, o que los individuos actuen por preferencias sociales tales como reciprocidad, empatía o por aversión a la desigualdad (V. Gouveia, 2011). Teniendo en cuenta las dos vertientes teóricas previas sobre el comportamiento humano (individualismo y colectivismo), este trabajo propone, mediante la simulación de una situación de fraude, determinar por qué algunos individuos, que de ahora en adelante llamaremos agentes, hacen uso de acciones y métodos coercitivos, y a qué se debe el éxito de estas acciones como mecanismo para prevenir la transgresión de su confianza.

Existen varios estudios que demuestran diferencias entre las emociones provocadas por un castigo de segunda y uno de tercera parte (Fehr & Fischbacher, 2004). El consenso es que el castigo de segunda parte se suele llevar a cabo con más frecuencia y con más fuerza en comparación con el castigo de tercera parte. Mientras el castigo de segunda parte puede representar un sentimiento vengativo promovido por la ira y venganza, el segundo se puede presentar por indignación moral frente a una injusticia. En ese sentido la venganza, el deseo de compensación y la indignación suelen ser las razones más frecuentes por las que se comete un castigo. A su vez, existe una relación directa entre las razones mencionadas y el papel que enfrenta el castigo como herramienta para hacer cumplir las expectativas de cualquier agente. Consideremos entonces, pertinente preguntar, ¿Cuál es el mecanismo que explica el éxito del castigo en el cumplimiento de las expectativas? Para otorgar una respuesta ante esta pregunta, es oportuno considerar el siguiente aspecto: el investigador debe contar con la

posibilidad de modelar y explicar la decisión de las víctimas y espectadores de sancionar a quienes han violado su confianza o alguna norma social, teniendo como hipótesis que existen estímulos emocionales, institucionales y económicos para el castigo con el propósito de recuperar su estado de bienestar inicial. En este artículo tomaremos el delito de fraude, una situación real que se presenta en la vida cotidiana de la sociedad, como un suceso que permite modelar la actuación de la conducta humana frente a la violación de la confianza. El fraude es un acto en el que un individuo, al que podemos designar como estafador, utiliza la confianza generada en su víctima para despojarla de dinero. La situación de fraude presenta similitudes con el escenario del Juego de Inversión de Berg et. al (1995), pues en ambos casos la confianza y el retorno de la inversión son el foco de atención.

El acto de fraude, a través de estafa, se plantea como una extensión del juego de inversión (Berg, Dickhaut, & McCabe, 1995) que involucra dos agentes activos y un agente pasivo[7]. El juego de confianza es secuencial con información perfecta para un solo agente y completa para ambos. Cabe anotar que, para la construcción de nuestro experimento, se introdujeron cambios el juego tradicional. La participación de un tercero, el uso de mecanismos de información como el *cheap talk* y la posibilidad de castigo. Las modificaciones del juego dieron lugar a cuatro tratamientos enmarcados en un diseño *between subjects*.

En el primer tratamiento, T1 (la línea base), el participante A1 recibe 100 unidades monetarias experimentales (UME) las cuales puede invertir en un proyecto que es administrado por el participante A2. A1, además, puede enviar un mensaje, no vinculante, a A2 con relación al porcentaje de la ganancia que espera recibir de vuelta. Cualquier cantidad que decida invertir es triplicada por el proyecto, por lo que A2 recibe tres veces la cantidad invertida por A1 y debe tomar la decisión de cuánto retornarle. Existe, además, en este juego, un participante A3 que en esta línea base, recibe un pago exógenamente determinado, y simplemente es un observador de toda la interacción entre A1 y A2. Los tratamientos 2 y 3, T2 y T3, son exactamente iguales a T1, pero A1 y A3, respectivamente, tienen la posibilidad de castigar a A2. Mientras que con T2 intentamos medir el efecto de la venganza por una contravención directa a la confianza, con T3 buscamos medir el efecto solidaridad vía un castigo proveniente por un tercero. El tratamiento 4, T4, es igual al T3, pero la estructura de pagos de A3 ya no es exógena, sino endógenamente determinada por la interacción entre A1 y A2. Con T4 pretendemos medir el efecto de la venganza por una acción indirecta, el cual denominamos efecto indignación.

En el experimento también realizamos una medición incentivada de las expectativas empíricas y normativas, de primer orden y segundo orden, entre los participantes (Bicchieri C. , 2017) con el objetivo de identificar el rol de las creencias colectivas en la toma de decisiones de inversión, retorno y castigo.

Los principales resultados del experimento señalan que el número de castigos ejercidos por los agentes externos es el mismo que los ejecutados por los agentes directamente implicados en la contravención, solo 2 de cada 10 personas que tenía la opción de castigar realmente lo hizo, esto se contrasta con la creencia de estas personas, ya que la mayoría de los individuos que participaron del experimento esperaban que, más de la mitad de personas que tuvieran

[7] Excepto en los tratamientos de Third Party Punishment

la posibilidad de castigar, lo hicieran cuando se presentara una contravención a la confianza.

Por otra parte, es evidente que la creencia sobre el castigo de tercero aumenta cuando realmente existe una herramienta coercitiva; es decir, solo cuando el mecanismo de castigo es real y tangible, la expectativa empírica de los miembros del grupo sobre el castigo a una contravención aumenta. Es importante señalar que la tecnología de castigo para este estudio permite a un participante eliminar todos los pagos de otro, lo que puede ser considerado para algunas personas como un elemento radical y, por esa razón, presentar una baja frecuencia.

El resto del documento se divide de la siguiente forma. En primer lugar, realizamos una presentación de la literatura relacionada a los juegos de confianza y los mecanismos extrínsecos e intrínsecos que pueden determinar sus resultados. En el apartado tres presentamos el diseño experimental y en el cuatro los resultados. En la sección cinco desarrollamos nuestra discusión y conclusiones.

2. Literatura Relacionada

2.1. La confianza, la honradez y la reciprocidad

Para este trabajo es fundamental encontrar en el *Investment Game* (Berg, Dickhaut, & McCabe, 1995) una forma de representar una contravención a las expectativas y confianza de los individuos. El juego de inversión es un claro ejemplo de dilema social existente, ya que la sola preocupación por el interés individual conduce a un resultado ineficiente desde el punto de vista colectivo. Los castigos de primera y segunda mano, y elementos de información son algunos mecanismos utilizados para reducir la tensión entre los intereses individuales y colectivos. La decisión de confiar es una decisión basada en tres consideraciones: i) el grado de seguridad asignado por parte del primer agente al segundo, concepto parroquialmente conocido como "ser digno de confianza", ii) la posibilidad de pérdidas potenciales en caso que el segundo agente no tenga ningún grado de reciprocidad positiva con el primer jugador, y iii) las ganancias potenciales en caso de que el segundo agente tenga algún grado de reciprocidad positiva y correspondencia a la confianza del primer agente.

Una contravención a la confianza genera sentimientos de frustración en los individuos directamente afectados e inclusive, entre los espectadores, quienes en algunos casos toman justicia a "mano propia"[8] para contrarrestar la pérdida social que deja el acto. Esta actitud individual de promover la *justicia a mano propia* es extremadamente riesgosa para la convivencia ciudadana, ya que puede contribuir a la conformación de grupos de autodefensa, que finalmente vuelven más complejo el panorama social y más difícil la solución al mismo.

En la actualidad se observa que el sistema judicial conserva profundos vacíos en su estructura funcional, posibilitando el funcionamiento de actividades ilícitas y la resignación de la comunidad ante medidas afectivas que las detengan. Según Donald Cressey (1961), el comportamiento de las personas que comenten infracciones a la confianza y los deseos de los demás individuos, como el fraude o estafa, pueden definirse como "violadores de la

[8] Es importante señalar que, en diferentes informes periodísticos de Latinoamérica, en su mayoría mexicanos, mencionan que los individuos que actúan de esta manera usualmente se autodenominan "justicieros".

confianza". Por esta razón se puede decir que el problema de fraude se agudiza cuando: i) no se castiga con severidad al estafador, ii) cuando surgen individuos que buscan justicia a través de alternativas ilegales, y aún más, iii) cuando se efectúa la detención del infractor pero no hay repercusiones judiciales, es decir, el infractor queda impune.

Fehr y Fischbacher (2004) aportaron a la comprensión de las normas sociales mediante el estudio de algunos mecanismos como la adición de un tercer agente con un opción de castigo en un juego del dictador (DG). Las normas son establecidas debido a la expectativa de que serán utilizadas como lineamiento para castigar las violaciones a las reglas de comportamiento. Sin embargo, Fehr y Fischbacher, al igual que Bendor y Swistak (2001), consideran que la existencia de sanciones a manos de terceros se constituye como la esencia de las normas sociales, ya que las estrategias de castigo de segundas partes no son sostenibles en el largo plazo, mientras que las estrategias que implican sanciones de terceros lo son.

Los conceptos de confianza y reciprocidad permiten comprender las causas de la acción colectiva como herramienta de solución a los problemas que afectan la dignidad de la comunidad, son principios fundamentales del tejido social, pues es la confianza entre ciudadanos lo que permite que una comunidad funcione. Enmarcar el análisis en estos conceptos significa trascender la idea de lo puramente racional y enfocarse en las definiciones morales y emocionales como estímulos para justificar la actuación del individuo. Es por ello que el tratamiento de los problemas de confianza y reciprocidad resulta difícil y poco convincente, ya que plantea un comportamiento humano que difiere, en muchas ocasiones, al evidenciado en la realidad.

Beneficiarse de la cooperación y confianza de terceros en detrimento de estos es una práctica lamentable (Fehr & Fischbacher, 2004). Boyd (2005) hace referencia al ***"castigo altruista",*** conocido término dentro de la literatura de las ciencias del comportamiento que puede dar solución a problemas individualistas como la falta de reciprocidad y el oportunismo. La acción del castigo altruista consiste en imponer sanciones a los abusadores (*free riders*), que explotan y violan la confianza de los demás, de manera que su comportamiento abusivo sea corregido y sean forzados a cooperar para evitar futuros castigos y futuras pérdidas. Cabe anotar que el castigo altruista supone un costo económico para el individuo que quiera ponerlo en acción.

2.2. Mecanismos Intrinsecos: Norma Social y Cheap Talk (Expectativas empiricas y normativas)

Los mecanismos intrinsecos están directamente relacionados con las motivaciones intrínsecas de determinada acción, es decir, con el interés por llevar a cabo una actividad por su satisfacción individual, más que por los posibles resultados/beneficios de la misma (Ryan, 2000) Son intrínsecos porque no requieren incentivos monetarios para su desarrollo. En este estudio hemos incorporado el estudio de las normas sociales y de las sugerencias no vinculantes como mecanismos intrínsecos que, si bien no tienen impactos monetarios, sí pueden modificar la conducta.

No existen comunidades sin normas sociales o estandares normativos que limiten el comportamiento. Más allá de conocer cuáles son las normas que regultan la accion humana, se ha prestado mucha atención a las condiciones bajo las que serán obedecidas esas normas. Debido a ello, la cuestión de las sanciones han sido importantes en la literatura de las ciencias

sociales. Una norma social es vista como elemento central para la produccion de orden o coordiancion social, y su investigacion se ha enfocado en las funciones que desempeñan y si lo hacen de manera eficiente. Incluso una norma puede determinar la maximazación del bienestar o la eliminación de externalidades, por ello las creencias, expectativas, el conocimiento de grupo y el conocimiento común se han convertido en conceptos para el desarrollo de una definición de norma social. Cristina Bicchieri (2006) define **las expectativas empíricas** como la creencia individual sobre el comportamiento efectivo, y las **expectativas normativas** como la creencia sobre el comportamiento aprobado por un grupo de personas, denominado red de referencia.

La capacidad de crear y de hacer cumplir las normas sociales es probablemente una de las caracteristicas distintivas de los humanos. Sin embargo, pese a su uso frecuente en muchos estudios, actualmente no existe un acuerdo colectivo sobre qué es exactamente una norma social. En años recientes encontramos aproximaciones relevantes a la definición del concepto hechas por varios estudiosos. Cristina Bicchieri (2017) define la norma social como las expectativas que cada sujeto tiene sobre lo que la red de referencia o grupo al que pertence hace y aprueba. Bicchieri (2006) también sostiene que cumplir las normas sociales es racional, debido a que cualquier incumplimiento puede dar cuenta de alguna violación a las preferencias y creencias de los demás actores. Las normas sociales son ampliamente invocadas para explicar datos que son dificiles de comprender con teorías alternativas de comportamiento (Bicchieri C. , 2017). Camerer y Fehr (2002), por ejemplo, sostienen que las desviaciones de las predicciones teóricas del juego basadas en intereses propios son interpretadas naturalmente como evidencia de las normas sociales. En el caso de castigos de tercera parte es probable que las motivaciones economicas estén ausentes. Como han demostrado Fehr y Fischbacher (2004), los observadores de terceros están dispuestos a castigar los infractores, pero incluso en este caso, no podemos afirmar el tipo de norma (moral o social) que guia el comportamiento.

Comprender la definición y el alcance de las normas sociales es importante para conocer la evolución de las relaciones humanas, pues tienen gran influencia sobre el debate entre el interés individual y el interés colectivo como mecanismos para tomar decisiones en la comunidad. Consideremos la situación en la que un individuo observador, no una victima directa de fraude, desee que el infractor sea castigado. Teniendo en cuenta que el sujeto obsevador es exógeno al fraude pero aún así ansía justicia, podemos concluir que la motivación de su conducta no está marcada por la venganza sino por el cumplimiento, o no, de lo que él cree que debe suceder. El sujeto simplemente no quiere impunidad, y la impunidad está relacionada con lo que el sujeto cree que ocurrirá y debe ocurrir.

En nuestro diseño experimental modificamos el juego de confianza involucrando un "mecanismo de información" no vinculante, o *Cheap Talk* (Bracht & Feltovich, 2008; Crawford & Sobel, 1982). El concepto de *Cheap Talk* hace referencia a la comunicación entre jugadores sin afectación directa en los beneficios o pagos finales. Una buena traducción de este mecanismo al español es "habladuría barata", es decir, un montón de palabras que no deberían tener ningún efecto directo sobre las decisiones de las personas. Un individuo podrá o no seguir las especificaciones entregadas a través del mecanismo de *cheap talk,* pero si decide desviarse de el, no habrá consecuencias directas sobre el mismo. El trabajo de Bracht

et. al. (2008) probaron varios mecanismos de información no coercitivos diseñados para aumentar la cooperación y la eficiencia en un juego de confianza. En este experimento se agrega una etapa previo al juego tradicional en la que el inversor recibe información a traves de un mensaje (*Cheap Talk*) sobre el comportamiento pasado del asignador. Ninguno de estos mecanismos altera los pagos finales de los participantes, pero se encontró que la cooperación y eficiencia aumentan sustancialmente cuando los inversionistas pueden observar el historial del asignador previo a la decision de asignar.

2.3.Mecanismos Extrínsecos: Castigo (First and Third Party)

Los mecanimos extrínsecos son aquellos elementos que para su funcionamiento necesitan de motivación externa, es decir, una recompensa o un castigo externo para desarrollar una actividad determinada. La segunda modificación del juego consiste en la introducción de la posibilidad de sanción a manos de la tercera parte. Se ha evidenciado en *ultimatum game* (UG) que los castigos por segunda parte son consistentemente más altos que los de terceros para aquellos sujetos que eligen compartir menos de la mitad de su dotación (Hoffman, 2008).

Nótese que nuestro diseño se desarrolla en un contexto de instituciones débiles, toda vez que no existe ninguna institución formal que castigue las contravenciones a la confianza. A diferencia del castigo suministrado por las autoridades policiales y judiciales, el castigo de terceros es ilegítimo, desde un punto de vista institucional. Está basado en la decisión personal de un individuo de corregir una falta percibida. En algunos lugares, el castigo de terceros, cuando se usa responsablemente, es una herramienta útil para hacer cumplir las normas sociales. Algunos estudios de la Universidad de Maryland (2013) sugieren que el castigo de terceros tiene mucha más probabilidad de evolucionar en contextos de alta restricción social y estructural porque en el largo plazo beneficia a toda la comunidad, incluyendo las víctimas de delitos. Sin embargo, puede que esta herramienta intensifique el conflicto y que la comunidad se vea envuelta en hechos sistemáticos de venganza.

El castigo puede definirse como una pena que se impone a la persona que ha incurrido en un delito o falta. Las sociedades a menudo acuden a la figura del castigo para sancionar las violaciones a las leyes establecidas. El castigo puede ser impuesto por una institución (Castigo centralizado), o por agentes individuales y (o) grupales que no gozan de legitimidad jurídica (castigo descentralizado) para sancionar la falta.

Según Clutton-Brock y Parker (1995), el castigo se configura como una acción que implica un costo para la persona castigada, pero también implica un costo para el castigador. Tal definición se refiere a la figura del castigo altruista o castigos de terceros (Third-Party Punishment). El tercero cuenta con la característica de ser un observador que presencia la transgresión sin verse afectado directamente por ella. Nikiforakis y Mitchell (2014) exploran dos caras de la misma moneda: el castigo y la recompensa. A través del juego del dictador, los autores encuentran que cuando se presentan simultáneamente las opciones de castigar o premiar las acciones de otro, la demanda por castigos costosos se reduce, pues se retiene la recompensa como una forma de castigo sin costo. De la misma manera, la demanda por recompensas costosas se reduce, pues la elección de no infligir un castigo resulta como una

forma de recompensa sin costo. La evidencia indica que tal comportamiento se debe a que la imposición del castigo (recompensa) corresponde a una forma de expresar la aprobación o desaprobación de los actos del otro más que un medio para alterar únicamente los pagos materiales de la persona analizada. Es así como los autores concluyen que restringir solo las opciones de castigo (recompensa) puede conducir a un incremento de la demanda por castigos costosos (recompensa).

Por su parte, Martjn Egas y Arno Riedl (2008)identifican los límites del castigo altruista para mantener la cooperación entre individuos que no están relacionados utilizando un juego de bienes públicos, de carácter repetido y con posibilidad de castigo. Los investigadores hallaron que la cooperación solo es mantenida si las condiciones para el castigo altruista son favorables (Bajo costo para el castigador; Alto impacto en el castigado), por lo cual, si las condiciones no son adecuadas, puede quedar un buen porcentaje de free-riders sin castigar. También agregan que el castigo altruista por sí solo no puede mantener la cooperación, se requieren de otros aspectos planteados por la literatura relevante del tema, por ejemplo, interacciones repetidas que permiten beneficios para el cooperador a través del altruismo recíproco y la construcción de reputación.

Wargo (2018) estudia el papel de la venganza y la aversión a la inequidad como determinantes del castigo utilizando el Juego de la Venganza (Revenge Game). El autor encuentra que una parte significativa de la muestra decide castigar cuando se encuentran en desventaja, pero también cuando el "tramposo" está escogiendo una situación general más justa. En ese sentido, Wargo concluye que la aversión a la inequidad explica una parte, pero no es responsable de la imposición del castigo, existe otro determinante: preferencias por la reciprocidad o, en otras palabras, venganza.

Carbonara y Fabbri (2017)exploran los efectos de la influencia social sobre la decisión de castigar de terceros, también utilizando el contexto del dictador que, a diferencia de un juego del dictador tradicional, no estudia las decisiones de envío sino las decisiones de apropiación de dinero. Los terceros rellenan un cuestionario donde lista sus decisiones de castigo ante diferentes situaciones (método estratégico). En un tratamiento (informacional) se les informa los castigos promedios de otros participantes. En el tratamiento normativo, se les dice que la tercera parte que sus decisiones de castigo serán juzgadas por 5 participantes seleccionados al azar. Los investigadores encuentran que cuando los terceros reciben información sobre el castigo hecho por los otros, revisan sus decisiones y reducen el castigo, por lo que estar de acuerdo con la mayoría parece ser más potente que su necesidad de tener razón (Influencia social informacional).

Meier et al (2016) analizan como la incidencia del crimen organizado puede generar disparidades en poblaciones italianas homogéneas a nivel étnico, religioso y lingüístico. Los autores implementan juegos experimentales en escuelas secundarias pertenecientes a dos sectores impactados de forma diferente por la mafia: Alta participación de la mafia vs Baja participación de la Mafia. Para analizar como varían las acciones de los estudiantes ante el castigo por parte de terceros, se implementa el clásico dilema del prisionero. Se encuentra que, ante la perspectiva del castigo, ambas áreas reportan un aumento de la cooperación. Por tanto, el crimen organizado no parece afectar negativamente la efectividad general de un mecanismo de aplicación de la norma. No obstante, el castigo no es capaz de resolver los

problemas de confianza y cooperación asociados con el crimen organizado, pues no eleva la cooperación en áreas de alta mafia al nivel presentado por la escuela de baja mafia. Adicionalmente, los estudiantes pertenecientes a las escuelas de la alta mafia mostraron un ligero favoritismo hacia el In-Group en el juego del dilema del prisionero sin castigo. Cuando se implementa el castigo, tal favoritismo aumenta. Cabe agregar que los estudiantes del sector de la baja mafia no muestran tal sesgo.

A pesar de que el castigo funciona en condiciones de laboratorio para aumentar la cooperación entre participantes, estos encuentran limitantes a la hora de explicar el mundo real. Guala (2012) aborda tal problemática planteando dos interpretaciones (Narrow y Wide) de los experimentos que involucran un castigo costoso. Bajo la interpretación del tipo Narrow, los experimentos que involucran el factor castigo son solo dispositivos útiles para medir propensiones psicológicas robustas (o en otras palabras, "preferencias sociales") en condiciones de laboratorio controladas. La interpretación del tipo Wide involucra la réplica de mecanismos que explican la cooperación tanto en el mundo real como en el laboratorio. Guala encuentra que solo la interpretación del tipo Narrow se encuentra sustentada con información experimental, mientras que la interpretación del tipo Wide requiere de evidencia de campo acerca de los mecanismos que permiten la cooperación fuera del laboratorio. No existe evidencia en la literatura antropológica de que el castigo costoso sea utilizado en pequeñas sociedades a excepción de la regulación de conflictos sexuales. Al contrario, existe vasta evidencia que pondera a la venganza como una de las mayores causas de disolución de los lazos sociales. La cooperación económica en pequeñas sociedades estudiadas por los antropólogos está usualmente soportada por mecanismos de castigo de muy bajo costo, como ostracismo y crítica verbal. Para concluir, Guala desafía la afirmación de que las preferencias sociales se expresan a través de castigos costosos que mantienen la cooperación en una vasta gama de situaciones que incluyen los escenarios de campo.

3. Diseño Experimental

3.1. Juego Experimental

El juego se desarrolló en dos fases: en la Fase I se llevó a cabo un juego de confianza y en la Fase II una licitación de expectativas empíricas y normativas. Este trabajo optó por decisiones *one shot* en un juego de confianza con posibilidad de castigo. El juego, propuesto originalmente por Berg et. al (1995), es un conocido ejemplo de un dilema social, en el que el interés propio de los agentes conduce a un equilibrio ineficiente desde el punto de vista social. Sin embargo, este trabajo busca poner a prueba el efecto de la participación de un tercero, las expectativas y la presencia de castigo en los niveles de cooperación entre los individuos.

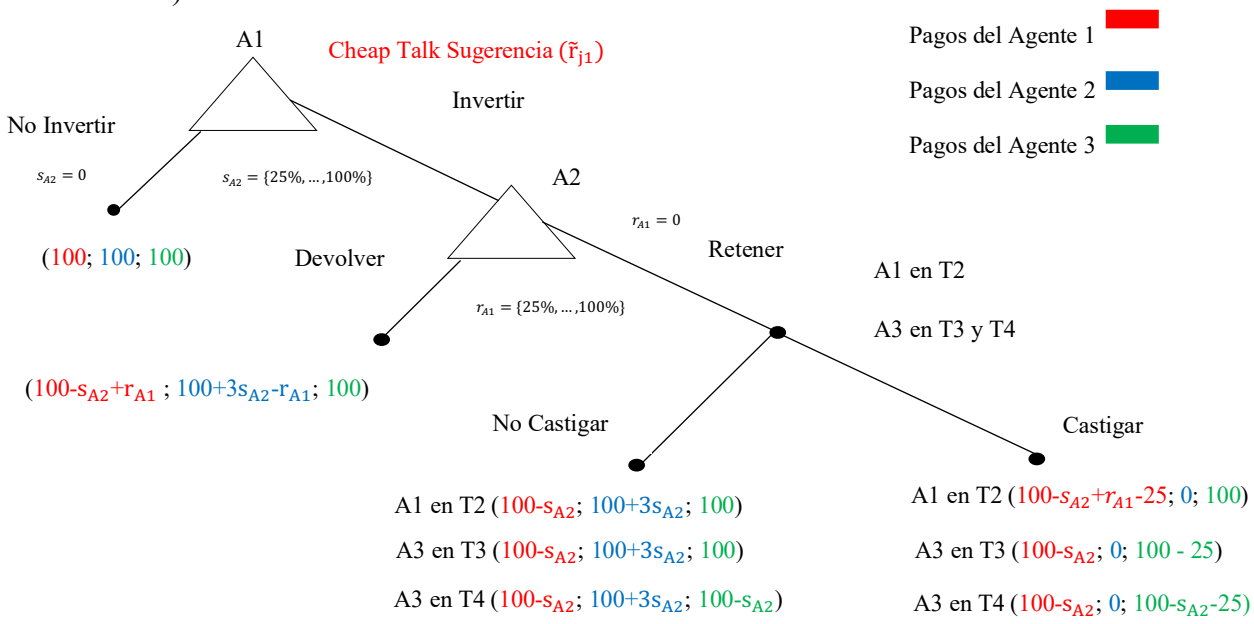
El juego de confianza es un juego secuencial con información perfecta para el primer agente y completa para ambos. El primer agente (*en adelante A1*), debe decidir si entrega o no parte de su dotación[9] (e) en un proyecto. El segundo sujeto (*en adelante A2*), quién administra el proyecto, recibe triplicada[10] la cantidad entrega por A1, que representa la rentabilidad de la

[9] La dotación viene endógenamente entregada por el experimentalista a ambos agentes por cantidades iguales.

[10] La cantidad que recibe A2 es el triple de lo que envía A1, esto para hacer el trabajo comparable con el de Berg et. al., 1995

inversión, y debe decidir cuánto retornarle. A diferencia de un juego simple de confianza, junto con el dinero entregado en el proyecto, A1 también envía una sugerencia, un contrato no vinculante, que indica lo que espera que A2 le retorne. Se incluye además la presencia de un tercer sujeto (*en adelante A3*), un espectador de la interacción entre A1 y A2, y que en algunos tratamientos cuenta con posibilidad de castigar a A2.

Para simplificar el análisis, la dotación inicial fue fijada en 100 Unidades Monetarias Experimentales (UME) entregadas a ambos agentes por igual, para evitar que la aversión a la desigualdad juegue algún rol. El espacio de estrategias de A1, su decisión de inversión en el proyecto administrado por A2 está representado por s_{A2}; la sugerencia de su expectativa de retorno se representa por el término $\widetilde{r_{A1}}$, y ambos son discretos, de tal manera que $s_{A2}\ y\ \widetilde{r_{A1}} \in \{0{,}25{,}50{,}75{,}100\ \text{UME}\}$. El espacio de estrategias de A2, su decisión de retorno también es discreto y consiste en retornar, r_{A1}, tal que $r_{A1} \in \{0\%, 25\%, 50\%, 75\%, 100\%\}$. Finalmente, A1 (o A3) pueden, eventualmente, tomar la decisión de castigar a A2 pagando $25C_i$ UME para destruirle completamente[11] sus pagos (para i=A1 o A3, según el tratamiento).



3.2.Tratamientos

Se condujeron cuatro tratamientos en el experimento con un protocolo *between-subjects*. El *timing* de cada uno y el resumen de los atributos básicos se encuentran representados en la tabla 1. En el primero, la línea base (BSL) consiste en el juego de confianza tradicional de tres agentes antes descrito y sin posibilidad de castigo. Con este tratamiento logramos establecer una línea base que permite mantener constante el efecto del observador. En el segundo, T2, transcurre igual que la línea base, pero incorpora, además, la posibilidad de que A1 castigue a A2 después de observar r_{j1}. En este tratamiento, el agente 3, A3, actúa como observador y sus pagos son fijos e iguales a 100 ECU. Dado que en T2 el castigo puede ser ejecutado por la víctima, nos permite estudiar la probabilidad de venganza y su impacto en

[11] Esto emula la decisión extrema de los "justicieros" de desaparecer a los infractores.

las decisiones de inversión y retorno. En el tercer tratamiento, T3, le hemos asignado a A3 la posibilidad de castigar a A2, asumiendo el costo del castigo, pero garantizando que sus pagos sean determinados exógenamente. Dado que en T3, A3 no se ve afectado por las decisiones de A1 y A2, la decisión de castigar nos permite estudiar la probabilidad de llevar a cabo una sanción imparcial, la cual hemos denominado *efecto solidaridad*. Finalmente, en el tratamiento cuatro, T4, A3 obtendrá lo mismo que obtenga A1 al finalizar su interacción con A2 y después de examinar sus ganancias, tendrá la posibilidad de castigar a A2, ya no como un observador imparcial, sino como un damnificado de las decisiones de ambos agentes. T4 captura lo que hemos denominado *efecto indignación*[12].

Tabla 1 Representación esquemática del Diseño Experimental y Tratamiento

	BSL	T2	T3	T4
Etapa I	A1 Entrega S + Sugerencia	A1 Entrega S + Sugerencia	A1 Entrega S + Sugerencia	A1 Entrega S + Sugerencia
Etapa II	A2 Retorna	A2 Retorna	A2 Retorna	A2 Retorna
Etapa III	_____	A1 Castiga	A3 Castiga	A3 Castiga

3.3.Obtención incentivada de creencias y normas

El comportamiento del grupo, a diferencia del comportamiento individual, se caracteriza por la similitud entre los rasgos y creencias del individuo frente a las del grupo. La cohesión, la tendencia a cooperar y a lograr objetivos comunes son atributos que permiten fortalecer cualquier vínculo del individuo con su red de referencia.

Un problema importante durante la evaluación de las expectativas sociales reside en que las expectativas comunicadas pueden no ser exactas, ya que los encuestados no tienen incentivo para pensar a profundidad en lo que creen que hacen y aprueban los demás. En este caso, podría haber una tendencia a proyectar las propias preferencias y creencias. Una solución que debería ser efectiva es incentivar la obtención de expectativas empíricas y normativas. Cuando la precisión en las respuestas promete una recompensa, los encuestados están motivados para hacer una conjetura más exacta (Bicchieri C. , 2017), por esa razón el diseño experimental implementó un bono de 8 UME a las personas cuyas suposiciones se desvían menos 5% de las del resto del grupo.

3.4.Hipótesis

Con los tratamientos ya definidos, las siguientes hipótesis se someterán a ensayo:

H1: La presencia de un castigo, C_{A1} en T2, C_{A3} en T3 y T4, genera incrementos en las cantidades retornadas por A2 ($\uparrow r_{A1}$) (Fehr & Gachter, 2000)

H2: Se espera que los castigos de A3 y A1 se vean afectados por las motivaciones a las que están expuestos:

$$C_{A3}^{t4} = C_{A1}^{t2} \rightarrow \text{El efecto venganza es el mismo para A1 y A3}$$

[12] Todos los tratamientos introducen el mecanismo de información *Cheap Talk* a través de la sugerencia.

$$\text{Si } C_{A3}^{t3} > C_{A1}^{t2} \rightarrow Predomina\ el \text{ Efecto solidaridad}$$
$$\text{Si } C_{A3}^{t3} < C_{A1}^{t2} \rightarrow Predomina \text{ el Efecto de venganza directa}$$
$$C_{A3}^{t3} > C_{A3}^{t4} \rightarrow Predomina\ el \text{ Efecto solidaridad}$$
$$C_{A3}^{t3} < C_{A3}^{t4} \rightarrow Predomina \text{ el Efecto indignación (venganza indirecta)}$$

H3: La expectativa empírica y la expectativa normativa que tienen los miembros del grupo sobre la frecuencia del castigo, es superior a la frecuencia real. (Fehr & Fischbacher, 2004)

3.5.Procedimiento

El experimento fue conducido y programado usando el software Z-tree (Fischbacher, 2007), en las instalaciones de la Universidad del Valle sede Meléndez en la ciudad de Cali. Los participantes fueron reclutados a través de mensajes al correo electrónico e invitados a participar de una preselección, a partir de la cual hacemos una asignación aleatoria.

Se condujeron cuatro sesiones conformadas por 30 individuos en cada tratamiento en un experimento one-shot, para un total de 30 observaciones independientes por rol y tratamiento, contando con un total de 120 personas. Una vez iniciada la sesión, los participantes, que se situaron en la sala de computo, recibieron instrucciones en la pantalla para ser leídas individualmente. La comprensión de las instrucciones fue puesta a prueba a través de un Quiz que los participantes respondieron previo al experimento. Los participantes que evidenciaron dificultades en la comprensión de las instrucciones obtuvieron una explicación adicional, privada y detallada para resolver inquietudes.

Posteriormente, se les informó a todos los sujetos sobre la dotación de cada agente, el tipo de cambio de unidades experimentales a pesos y también se explicó la representación del juego al comienzo del experimento, permitiendo a los participantes conocer todas las estrategias del juego para cada agente. Los sujetos interactuaron de forma anónima y nunca fueron informados acerca de las identidades de los otros agentes. Así mismo, se aclaró que los agentes podían incurrir en una pérdida de unidades monetarias por sanciones y que dicha perdida tenía que ser descontada de los pagos al final de la actividad. Después de que todas las decisiones fueron tomadas, los sujetos eran informados sobre el resultado de sus pagos.

3.6.Aspectos Éticos

Aunque la investigación requirió de la interacción con seres humanos, las temáticas que se abordaron en busca de determinar los mecanismos que permiten que el castigo sea un éxito en el cumplimiento de las expectativas de los agentes no interfirieron en su vida privada o íntima.

Con respecto a los lineamientos éticos del estudio en ciencias sociales, se tuvieron en cuenta dos principios esenciales en la interpretación de los derechos de las personas. El primero se refiere a respetar las decisiones de los sujetos y protegerlos de daños a través del anonimato, y el segundo, dar un trato a las personas como agentes autónomos, por tanto, la investigación respetó la voluntariedad de participar.

Como la participación fue voluntaria se hizo necesario el diligenciamiento del "consentimiento informado" del sujeto de investigación. Esto garantizó la autonomia y libertad de los participantes, la relevancia del proyecto y controló los riesgos en que las

personas puedan incurrir al participar.

4. Resultados

A continuación, se analizarán los incentivos conductuales y económicos que conducen a individuos, víctimas de vulneración de su confianza o testigos de dicha vulneración, a administrar justicia por su propia cuenta.

Tabla 2 Promedio por variable y por tratamiento

Tratamiento	Número de Observaciones	Media de inversiones	Media de Retornos	Media Sugerencia	Promedio Ganancias ($)
BSL	30	.50 (.1054)	.25^{AB} (.0912)	.425 (.075)	14.877 (951)
T2	30	.55 (.0971)	.50^{A} (0)	.60 (.0666)	14.233 (595)
T3	30	.40 (.0666)	.425 (.0989)	.475 (.0786)	12.930 (625)
T4	30	.55 (.0971)	.50^{B} (.0745)	.475 (.0692)	15.524 (800)
Total	120				

Fuente: Elaboración propia de los autores con base en información arrojada por Z-tree. La desviación estándar se encuentra entre paréntesis. Se contrastaron los promedios de la cantidad invertida por tratamiento con una prueba Wilcoxon – Mann – Whitney encontrando que no existen diferencias significativas entre tratamientos por esa razón no hay superíndices en los datos. Para la comparación del promedio de retornos entre tratamientos se usó el test Wilcoxon – Mann – Whitney (Ranksum) encontrando diferencias significativas, los datos cotejados tienen el mismo superíndice por ejemplo BSL y T2, A, que da cuenta de un p-valor de 0.05; y BSL y T4 tienen el superíndice B, que representa un p-valor de. 0.05. Aquellos datos que no tienen superíndices es porque la diferencia estadística no es significativa.

4.1. Actos de Confianza y Reciprocidad

En la tabla 2 se registró la cantidad enviada por los individuos A1 en los cuatro tratamientos. Para la BSL, en la que no existe posibilidad de castigo y solamente A1 hace una sugerencia de lo que él cree debe ser devuelto por A2, el 60% de los Agentes 1 hace una entrega entre el 25% y el 50% de su dotación inicial, para un promedio de envío del 50% de la dotación inicial. En el T2 y el T4, el nivel de transferencia promedio aumentó en cinco puntos porcentuales, y en el tratamiento T3, los envíos fueron 10 puntos porcentuales menores que la línea base. Las diferencias entre los envíos se contrastaron con una prueba Wilcoxon – Mann – Whitney, la cual arrojó que no existen diferencias significativas entre tratamientos sobre la cantidad invertida por parte de los A1. Por lo tanto, no se puede asegurar que las instituciones sancionatorias promuevan la confianza de los agentes.

Por otra parte, el Agente 1 pudo no enviar nada de dinero a A2 ya que reconoce que éste último no tiene incentivos racionales para retornar los réditos de la inversión; sin embargo, en los cuatro tratamientos del experimento se registraron cantidades positivas de entrega[13]. Esto permite deducir que existe algún grado de confianza y una expectativa de retorno para su contraparte.

[13] Se comparó usando un Sign Rank Test si las Entregar = 0 vs Entregas > 0 obteniendo que: BSL (0.0105 $p<0.05$); T2 (0.0072 $p<0.01$); T3 (0.0051 $p<0.01$) y T4 (0.0099 $p<0.01$).

La dotación inicial de todos los participantes era la misma y de dominio público, así que si A1 ofrecía todo su *capital* (100 UME), el Agente 2 podía inferir rápidamente que, si no retornaba algún dinero, A1 no tendría la manera de pagar para castigarlo, pero si A1 había dejado suficientes recursos para ejecutar el castigo, A2 debería tener más incentivos para reciprocar. El porcentaje enviado por los A1 en los tratamientos T2 y T4 se concentra mayoritariamente entre 25% y 75% UME, lo cual indica que un 80% (70%) de los A1 (A3) en el tratamiento T2 (T4) contaba con los recursos necesario para pagar por el castigo.

Una vez demostrado que los individuos hicieron entregas positivas al segundo agente, resta por examinar el comportamiento de A2. En la cuarta columna de la tabla 2 se observa que la BSL fue el tratamiento en el cual el promedio de la cantidad retornada fue de una cuarta parte de lo que recibió A2, además el 50% de los A2 retornaron 0 UME. En los tratamientos que existe presencia de castigo, A2 tuvo, usualmente, inclinaciones por retornar entre el 40% y el 50% del dinero que le llegaba.

Para conocer si existen diferencias significativas entre los promedios de retorno se ejecutó un test de suma de rangos de Wilcoxon, obteniendo diferencias significativas en el aumento de las cantidades retornadas entre la línea Base (BSL), el tratamiento 2 (T2) y el tratamiento 4 (T4): Tanto en T2 como en T4 hay presencia de castigo; en el primero a manos de A1 y en el segundo a manos de A3. Sin embargo, hay una similitud que parte del hecho que el Agente 3 en T4 hereda el mismo nivel de pagos de A1. En consecuencia, A2 parece estar respondiendo recíprocamente cuando percibe la posibilidad de que se ejecute un castigo motivado por la *venganza.* La falta de significancia estadística entre los retornos de T3 y T4, y entre T4 y la BSL, permite inferir que A2 no parece responder a la posibilidad de que haya justicia de mano propia motivada por la *solidaridad* ni por la *indignación.*

4.2.Efectos del Castigo y las Creencias

El castigo es un mecanismo viable para aplicar la norma de aversión a la desigualdad o de la reciprocidad y compensar al individuo que ha sido víctima de alguna injusticia. Con la presencia de este elemento, se espera que los A2 cuenten con las motivaciones intrínsecas necesarias para cumplir con las expectativas de los Agentes 1. Es importante mencionar que la expectativa empírica de A1 sobre el castigo de A2 (es decir el porcentaje de personas que A1 creía que iban a castigar a A2) es incluso más alta que el castigo real, pues los individuos respondieron en la BSL que, en promedio, alrededor del 48.16% de los que tenían la opción de castigar, lo iban a hacer. Este porcentaje se incrementó, como era de esperarse en los demás tratamientos en los cuales la posibilidad de castigar era efectiva.

En la misma línea de Fehr & Fischbaher (2004), encontramos que, de 30 individuos, 24 dejaron el dinero suficiente para pagar por el derecho a castigar, pero realmente solo 6 lo hicieron. Es decir, solo el 25% de todos los individuos del experimento que tuvieron los recursos disponibles, castigaron y dejaron A2 sin dinero. En los tratamientos T2, T3 y T4, tan sólo el 20% de los participantes ejecutaron el castigo, pero a la vez creían que entre el 50% y el 70% de los demás individuos aplicarían la sanción. Es decir, hemos encontrado evidencia empírica de un fenómeno de ignorancia pluralística que reside en el hecho de que las creencias de lo que los demás hacen y aprueban, distan dramáticamente de lo que en

realidad ocurre (Bicchieri, 2017).

Somos conscientes de que la tecnología del castigo que hemos impuesto en el experimento es severa y que los agentes pueden sentirse moralmente impedidos para utilizarla; esto puede explicar las bajas tasas de castigo. Sin embargo, no explican las dimensiones de las expectativas empíricas y normativas. Por lo tanto, consideramos que la tecnología del castigo no está afectando la presencia de la ignorancia pluralística.

Tabla 3 Proporción de las Expectativas de Castigo y Castigos Efectivos

Variables	Castigo	Expectativa Empírica	Expectativa Normativa
BSL	N/A	48.16%^{ABC}	51.16%
T2	20%	64.2%^{ADE}	59.66%
T3	20%	66.5%^{BDF}	61.53%
T4	20%	59.06%^{CEF}	67.86%

Nota 1: los datos cotejados tienen el mismo superíndice; por ejemplo BSL y T2, A, BSL y T3 tienen el superíndice B, y así sucesivamente. Aquellos datos que no tienen superíndices son porque la diferencia estadística no es significativa.
Nota 1: Para la comparación del promedio de la expectativas empírica de castigo entre tratamientos se usó el test Wilcoxon – Mann – Whitney (Ranksum) encontrando diferencias significativas, los datos cotejados para la BSL y T2 A. P-valor = 0.000; BSL y T3 B. P-valor = 0.000; BSL y T4 C. P-valor = 0.000; T2 y T3 D. P-valor = 0.000; T2 y T4 E. P-valor = 0.000 y T3 y T4 F. P-valor = 0.000

En la tabla 3 se evidencia que existe una diferencia significativa sobre la expectativa empírica de castigo entre los sujetos que se encuentran en la línea base y el tratamiento 2, 3 y 4. La expectativa empírica de todos los individuos que participaron en el experimento, independientemente que se sean afectados directa o indirectamente por el castigo, sobre castigar el comportamiento abusivo del A2 es mucho mayor en los tratamientos donde existe la herramienta de coerción que en la BSL siendo T4 el tratamiento con mayor expectativa. Además, es importante señalar que la expectativa normativa no cambia entre tratamientos.

5. Discusión y conclusiones

The Trust Game es un experimento diseñado para revelar los mecanismos que favorecen la confianza y la reciprocidad en situaciones de no cooperación. Es de gran interés y valía descubrir mecanismos que permitan incrementar la cooperación y, de esta manera, incrementar la eficacia. Algunos de los mecanismos previamente estudiados consisten en la implantación de un tercero con poder coercitivo, cuyo propósito es incentivar el cumplimiento de acuerdos no obligatorios entre los agentes.

En este trabajo se abordaron temas como la sanción, las violaciones a la confianza y las normas sociales como conceptos estructurales para estudiar los castigos de terceros a infractores de confianza y comprender el fenómeno de la "justicia a mano propia" en países latinoamericanos. Los resultados obtenidos sugieren que, la posibilidad de castigo se use o no, funciona como un mecanismo para mejorar la reciprocidad, pues solo así el *free rider* tiene incentivos para no desviarse de las expectativas de A1. No obstante, la posibilidad de castigar no parece tener ningún efecto sobre las inversiones de los jugadores, ni sobre sus sugerencias de retorno.

De manera interesante, encontramos que nuestros tratamientos no tienen ningún efecto sobre la probabilidad de castigo, pero sí sobre las cantidades retornadas. Creemos que la severidad de la tecnología de la sanción desalienta a los jugadores a castigar, lo cual puede estar capando nuestro efecto tratamiento. No obstante, el temor a dicha severidad promueve la reciprocidad, lo cual redunda en diferencias significativas entre los retornos con y sin castigo.

En la misma línea del resultado anterior, se observa que la sugerencia, per sé, no tiene ningún efecto sobre la reciprocidad, y por lo tanto debe estar acompañada de la expectativa de una sanción; por eso, en la línea base los retornos son significativamente inferiores a los de los demás tratamientos, excepto T3.

Nuestro diseño experimental permite distinguir entre dos motivaciones que pueden estar afectando la decisión de castigar: *la solidaridad, la indignación y la venganza.* Al comparar los tratamientos T3 y T4 con respecto a la línea base, podemos aislar perfectamente estas motivaciones. Encontramos que los retornos de los A2 tienden a ser significativamente más altos con respecto a la BSL en T4, pero no hay diferencias significativas en T3. Esto nos indica que el efecto venganza es un motivador más intenso para el comportamiento recíproco, y que los sujetos le asignan una baja probabilidad de ocurrencia a las sanciones solidarias.

Nuestros resultados también sugieren que entre una gran proporción de nuestros participantes hay una fuerte creencia en que las transgresiones a la confianza serán sancionadas por el sujeto que tenga la posibilidad de hacerlo. Sin embargo, observamos que el porcentaje de personas que se cree que aprueban y llevarán a cabo la sanción, está entre 30 y 50 puntos porcentuales por encima del porcentaje efectivo de personas que castigan. Este fenómeno aporta evidencia empírica de ignorancia pluralística en la cual las expectativas de lo que ocurrirá (y de lo que los demás consideran que es lo correcto) se encuentran desfasadas con respecto a los hechos. Este desfase juega a favor de la eficiencia pues, la creencia de que se llevarán a cabo los castigos promueve la reciprocidad sin necesidad de que tengan que hacerse efectivos.

Finalmente, este trabajo muestra una disonancia con respecto a la consideración de la economía ortodoxa sobre la presunción de que el hombre es netamente racional, egoísta y que confronta información perfecta y completa. El experimento llevado a cabo evidencia que la naturaleza humana, si bien racional, también está basada en las emociones e indudablemente se comprueba que la pertenencia a un colectivo ocasiona modificaciones a las preferencias y acciones de un sujeto.

Referencias

Bendor, J., & Swistak, P. (2001). The Evolution of Norms. *American Journal of Sociology*, 1493-1545.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity and social history. *Games and Economic Behavior*, 122-142.

Bicchieri, C. (2006). The Grammar of Society. The Nature and Dynamics of Social Norms. *Cambridge University Press*.

Bicchieri, C. (2017). *Norms in the Wild: How diagnose. measure, and change social norms.* New York: Oxford University Press.

Boyd, R. H. (2005). *The Evolution of Altruistic Punishment.* MIT Press: Ed. H. Gintis. Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life.

Bracht, J., & Feltovich, N. (2008). An Experimental study of information mechanisms in the trust game: effects of observation and cheap talk. *University of Aberdeen Business School*, 22.

Camerer, C., & Fehr. (2002). Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists. *Working Paper No. 97, January. Institute for Empirical Research in Economics, University of Zurich*.

Clutton-Brock, T., & Parker, G. (1995). Punishment in Animal Societies. *Nature. 373.*, 209-16.

Crawford, V., & Sobel, J. (1982). Strategic Information Transmission . *Econometrica Vol 50 No 6*, 1431-1451.

Cressey, D. (1961). *The prison: Studies in institutional organization and change.* New York: Holt, Rinehart and Winston.

Egas, M., & Riedl, A. (2008). The Economics of Altruistic Punishment and the Maintenance of Cooperation . *Proceedings Biological Sciences*, 10.

Fabbri, M., & Carbonara, E. (2017). Social influence on Third-Party Punishment: An Experiment. *Journal of Economic Psychology* , 204-230.

Fehr, E., & Fischbacher, U. (2004). Third-party Punishment and Social Norms. *Evolution and Human Behavior* , 171-178.

Fehr, E., & Gachter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *The American Economic Review Vol. 90, No. 4*, 980-994.

Fischbacher, U. (2007). Z-tree: Zurich Toolbox for Readymade Economic Experiments. *Experimental Economics*, 171-178.

Guala, F. (2012). Reciprocity: Weak or Strong? What punishment experiments do (and do not) demonstrate . *The Behavioral and Brain Sciences, 35(1)*, 1-15.

Hoffman, E. M. (2008). Reciprocity in Ultimatum and Dictator Games: An Introduction. *Handbook of Experimental Economics Results* , 411 - 416.

Maryland, P. U. (2013). Game theory used to explain evolution of "Third Party Punishment" .

Meier, S., Pierce, L., Vaccaro, A., & La Cara, B. (2016). Trust and In-Group Favoritism in a Culture of Crime. *Journal of Economic Behavior & Organization* .

Nikiforakis, N., & Mitchell, H. (2014). Mixing the Carrots with the Sticks: Third Party Punishment and Reward . *Experimental Economics*, 17.

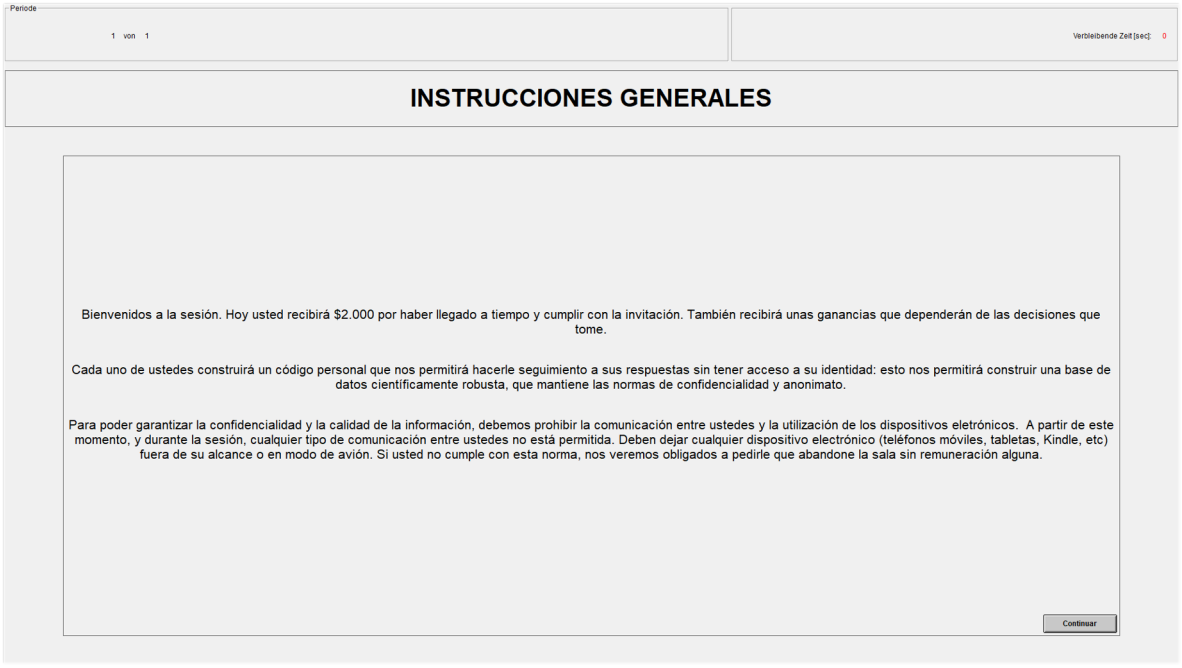
Ryan, R. M. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology, 25(1)*, 54-67.

V. Gouveia, T. M. (2011). Individualism-Collectivism as predictors of prejudice toward Gypsies in Spain. *Interamerican Journal of Psychology* , 45 (2), pp. 223-234.

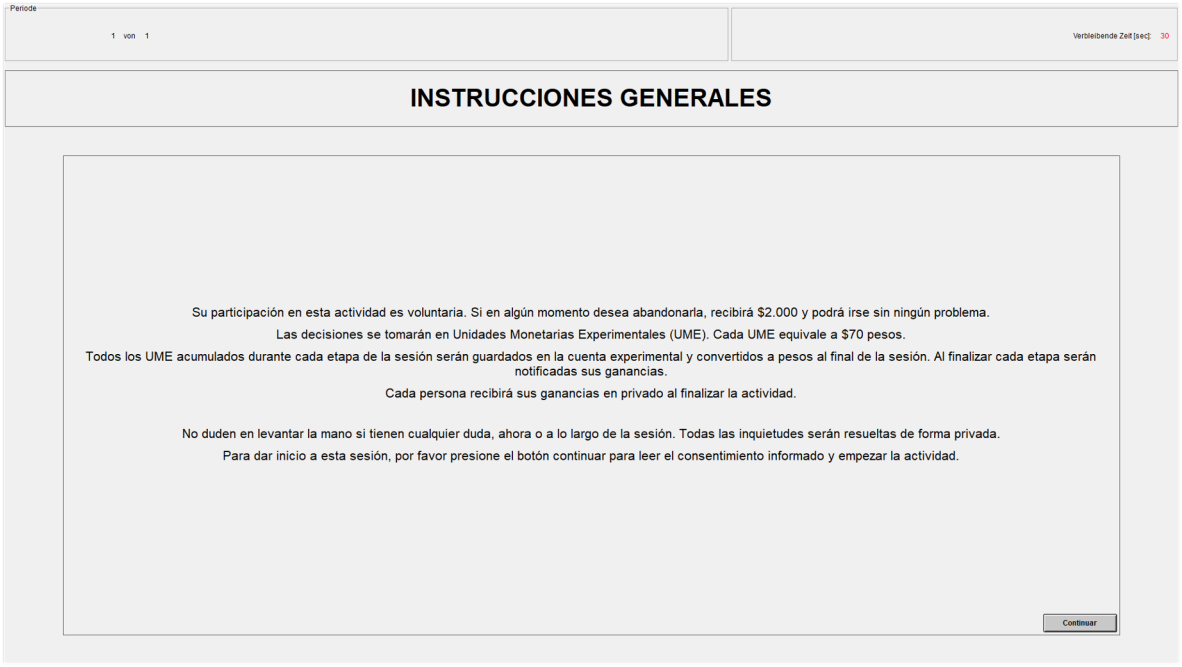
Wargo, D. T. (2018). Punishment As Revenge, Not Only For Inequity Aversion. *DETU Working Papers 1803, Department of Economics, Temple University*, 10.

Apéndice: Protocolo para un Juego de Confianza con Posibilidad de Castigo

Anexo 1. Instrucciones Generales del Juego Parte I

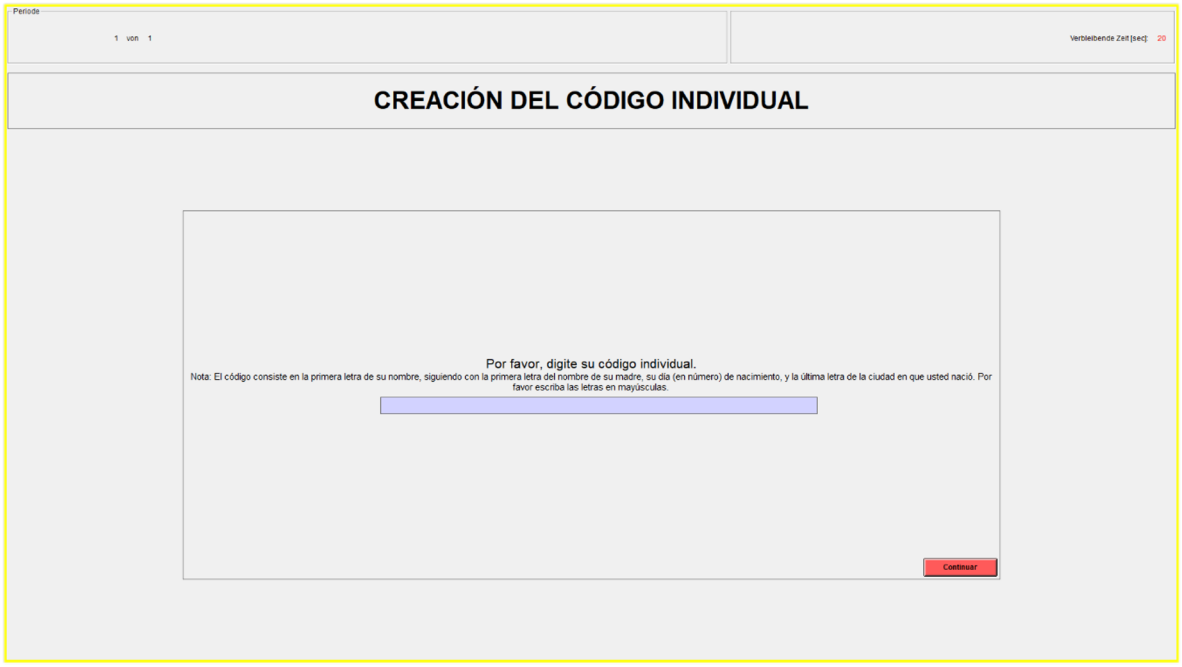


Anexo 2. Instrucciones Generales del Juego Parte II

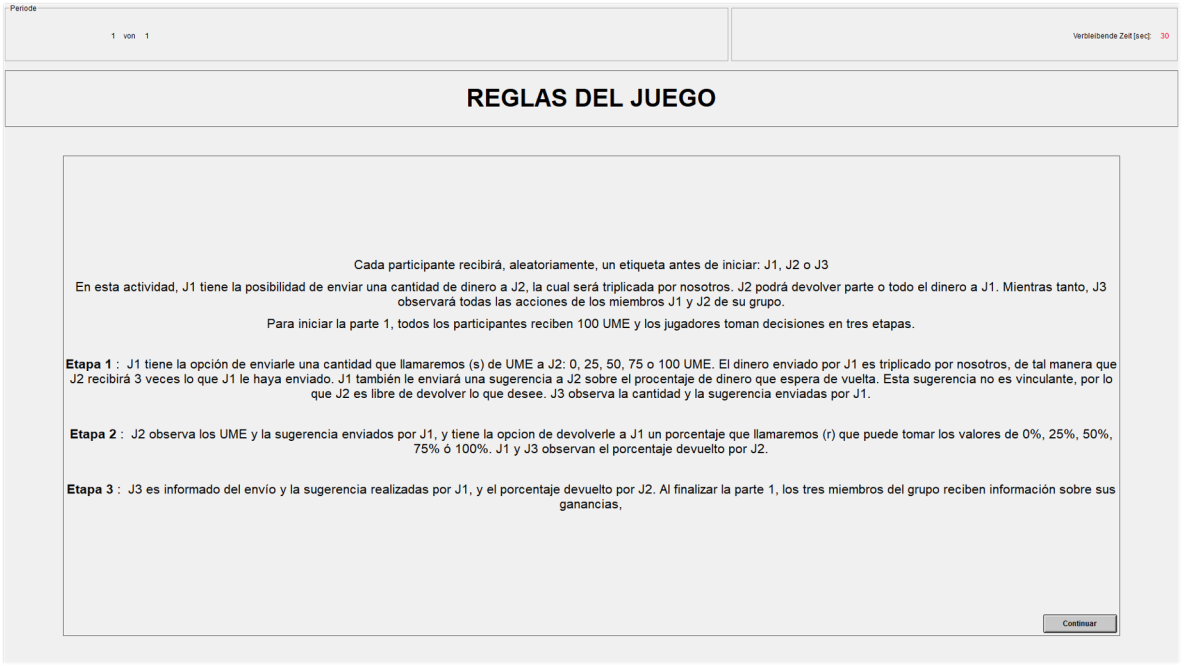


Anexo 3. Consentimiento Informado

Anexo 4. Creación del Código Individual



Anexo 5. Reglas del Juego



Anexo 6. Quiz de la Línea Base

Periode

1 von 1

Verbleibende Zeit [sec]: 40

QUIZ

Por favor, calcule las ganancias de los tres jugadores con base en el ejemplo que se muestra a continuación.
Sus respuestas no tendrán ninguna consecuencia en sus ganacias, simplemente queremos verificar que ha comprendido las instrucciones.
Si tiene alguna pregunta, no dude en levantar la mano. Su inquietud será atendida en privado.

J1 le envía 70 UME a J2 y le sugiere le regrese 100 UME.
J2 Recibe 210 UME (70x3=210), y decide regresar 0 UME a J1.
J3 Observa las acciones de J1 y J2.

Dotación Inicial de J1.

Cantidad que envía J1 a J2 (sin multiplicar):.

Lo que recibe J1 de J2:

Total Ganancias de J1:

Dotación Inicial de J2:

Lo que recibe J2 de J1 (x3):

Lo que retorna a J1:

Total de Ganancias de J2:

Total de Ganancias de J3:

Continuar