Universidad de Granada
Departamento de Métodos Cuantitativos
para la Economía y la Empresa

TESIS DOCTORAL
Análisis de las Decisiones Individuales en Contextos de Negociación. Aportes Teórico y Experimental

Doctorando
Luis Alejandro Palacio García

Directores
Teresa María García Muñoz
Juan Antonio Lacomba Arias
Francisco Miguel Lagos García

Editor: Editorial de la Universidad de Granada
Autor: Luis Alejandro Palacio García
D.L.: GR 516-2012
ISBN: 978-84-694-6003-0

La memoria de tesis titulada **“Análisis de las Decisiones Individuales en Contextos de Negociación. Aportes Teórico y Experimental”**, que presenta el doctorando *D. Luis Alejandro Palacio García* para optar al título de Doctor en Economía, ha sido realizada en el Departamento de Métodos Cuantitativos para la Economía y la Empresa, dentro del programa de doctorado Empirical Economics (E2), bajo la dirección de los Doctores *Dña. Teresa María García Muñoz*, *D. Juan Antonio Lacomba Arias* y *D. Francisco Miguel Lagos García.*

Granada, mayo de 2011

Doctorando

Fdo: D. Luis Alejando Palacio Gracía

Directores

Fdo: Dña. Teresa María García Muñoz   Fdo: D. Juan Antonio Lacomba Arias

Fdo: D. Francisco Miguel Lagos García.

Agradecimientos

Esta tesis no sería posible sin el trabajo de mis directores, Teresa, Juan Antonio y Francisco. Ellos me mostraron el camino de la investigación en economía experimental, desde redactar a prueba de fallos las instrucciones, pasando por diseñar y programar, hasta analizar los resultados y presentarlos de forma profesional. Pero sobre todas las cosas debo agradecer que siempre me hayan corregido con respeto y amabilidad, incluso cuando yo insistía en remar contra la corriente.

Agradezco a los proyectos del Ministerio de Educación y Ciencia (SEJ2009-11117/ECON) y de la Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía (P07-SEJ-3261) el apoyo económico para la realización de los experimentos de esta tesis doctoral. De igual forma, fue fundamental para mí contar con la beca de formación de personal docente e investigador predoctoral en áreas consideradas deficitarias por necesidades docentes, otorgada por la Secretaría General de Universidades, Investigación y Tecnología de la Junta de Andalucía.

Más que agradecer con palabras, tengo un compromiso y una responsabilidad inmensa con a la Universidad Industrial de Santander. Sin la comisión de estudios hubiese sido prácticamente imposible realizar el doctorado. En particular, debo agradecer el apoyo y la confianza depositada por Susana Valdivieso, una gran profesora, jefe y amiga. Espero contribuir con mi trabajo a construir la escuela que ella siempre soñó.

Tengo que hacer una mención especial para Manuel, él ha logrado demostrar que el alumno debe superar al maestro. Quién podía pensar que un proyecto escrito en una hoja arrugada nos convertiría en todos unos doctores. Y eso que siempre encontró una excusa ingeniosa para no leerse el libro de Milgrom. Y ni que decir del libro negro de Elster.

Para mi madre y mi padre el agradecimiento es obvio. Sin su ADN no estaría yo aquí. Pero además, ellos fueron mis primeros maestros en el arte de la negociación. Me llevó mucho tiempo demostrarlo científicamente, pero en esencia ellos tenían la clave: sensibilidad y compromiso. A mi madre le debo la sensibilidad para ponerme en la piel del otro, y a mi padre le debo el compromiso hacia la sociedad. Básicamente la metodología consistía en ensayo y error. Cuando me sentía perdido y no entendía por qué el mundo es como es, ellos no me decían lo que tenía que hacer, dejaban

que me enfrentará a la vida, siempre pendientes para apoyarme en los momentos difíciles. Bueno, a decir verdad mi padre si me decía lo que tenía que hacer, pero yo terminaba haciendo lo contrario.

Para Alexandra, el agradecimiento es infinito, si no fuera por ella no sería quien soy, ni tendría lo que tengo. Es más, no tiene sentido hablar de dos personas, compañeras y colegas, es mejor hablar de Palacio Cortes Compañía iLimitada, ambiente familiar, atendido por sus propietarios. Dadas las buenas expectativas, próximamente estaremos ampliando la nómina. Con el permiso de Andrea, no le doy más vueltas, uno lo mío y lo tuyo, porque sin lugar a dudas, mi vida favorita está pasando aquí y ahorita.

Definitivamente, no sé si es más difícil redactar la tesis o escribir los agradecimientos. Si intento hacer lo segundo con la misma rigurosidad de lo primero, pues simplemente me estaría enfrentando a una misión imposible. Me gustan los retos, pero no me veo capaz de enumerar a ciencia cierta las personas que participaron leyendo, opinando, corrigiendo, teniéndo paciencia, dando ánimo, acompañandome en los momentos de crisis y en los momentos de felicidad. Esta tesis es un gran logro personal, como consecuencia de un esfuerzo colectivo. Y como todo colectivo, siempre va lleno.

"The essence of these tactics is some voluntary but irreversible sacrifice of freedom of choice. They rest on the paradox that the power to constrain an adversary may depend on the power to bind oneself; that, in bargaining, weakness is often strength, freedom may be freedom to capitulate, and to burn bridges behind one may suffice to undo an opponent."

Thomas C. Schelling

"Conceptualmente lo que debemos diferenciar con claridad son las reglas y los jugadores. El propósito de las reglas es definir la forma en que el juego se desarrolla. Pero el objetivo del equipo dentro del conjunto de reglas es ganar el juego a través de una combinación de aptitudes, estrategia y coordinación; mediante intervenciones limpias y a veces sucias."

Douglass North

"De donde vengo yo, la cosa no es fácil pero siempre igual sobrevivimos. Vengo yo, de tanto luchar siempre con la nuestra nos salimos. Vengo yo, de aquí se habla mal pero todo está mucho mejor. Vengo yo, tenemos la lluvia, el frio, el calor."

ChocQuibTown

Índice general

Índice de figuras

Índice de cuadros

Capítulo 1

Introducción

1.1. El Poder de Negociación de las Amenazas

Conflicto y negociación están estrechamente relacionados, como se puede observar en una amplia variedad de situaciones de la vida diaria. Los gobiernos se enfrentan por controlar un territorio, los sindicatos se van a la huelga, las disputas legales terminan en los tribunales, las firmas compiten en agresivas guerras de precios, las parejas discuten sobre la distribución de tareas. Cuando no se puede llegar a un acuerdo, la colisión de intereses puede llevar a las partes en conflicto al peor escenario posible. Sin embargo, el desacuerdo puede ser usado para ganar poder de negociación por medio de amenazas. Una de las partes puede persuadir a su oponente a cooperar si le presenta el problema como una decisión entre solo dos alternativas: un resultado favorable para quien usa dicha estrategia, o en otro caso se llegaría al mutuamente indeseable desacuerdo.

El poder de negociación hace referencia a las habilidades relativas que tienen los agentes para ejercer influencia unos sobre otros. Dentro de este contexto, la amenaza es un mensaje que busca el beneficio propio, diseñada para hacer que el otro vea las consecuencias de sus acciones. Implica forzar al adversario de forma agresiva, pero en caso de ser exitosa disuadiendo el conflicto, podría llegar a ser mutuamente beneficiosa. Esto es precisamente lo que hace un gobierno cuando establece que está preparado para responder con armas atómicas ante cualquier provocación militar, o cuando un sindicato anuncia que está dispuesto a soportar una huelga costosa si no se cumple con sus demandas. Claramente, en estos ejemplos la comunicación está siendo utilizada con un propósito táctico, tomar ventaja en la negociación.

El objetivo de esta investigación es analizar el uso estratégico de la comunicación en una negociación bilateral. En particular, el objeto de estudio será el uso de las amenazas para ganar poder de negociación. Por amenaza se entenderá lo siguiente: 1) es un mensaje sobre la intención de ejecutar un plan de acción, 2) debe ser rentable para quien la formula, y 3) el mensaje lleva implícitamente a cuestionar su credibilidad. Las dos primeras características se han identificado en los ejemplos presentados, por lo tanto, es importante explicar en detalle en qué consiste el problema de la credibilidad.

Una amenaza busca cambiar las creencias que tiene el rival sobre las respuestas posibles a sus acciones. Sin embargo, el simple hecho de enviar un mensaje no es suficiente ante un rival que piensa estratégicamente, dado que no siempre las palabras coinciden con las acciones. En teoría de juegos, la credibilidad está asociada al concepto de Equilibrio Perfecto en Subjuegos, es decir, un mensaje es creíble si quien lo envía nunca tiene incentivos a desviarse del plan de acción. Desafortunadamente, esta definición es muy restrictiva, y en la práctica implicaría que cuando un mensaje es creíble, no podría ser visto como una amenaza. Para ilustrar esta idea podemos considerar nuevamente los anuncios de una nación o un sindicato respectivamente: "si somos atacados, entonces nos defenderemos" o "si se incrementan los salarios, entonces estaremos muy felices y agradecidos". En estos anuncios la credibilidad no está en duda, pero difícilmente podremos decir que son una amenaza.

Las amenazas que se analizarán en esta investigación implican que el emisor del mensaje debe anunciar un plan de acción tal que no tiene incentivos a llevar a cabo al menos una acción, como lo es llegar a una costosa confrontación. Quien envía el mensaje no desea el conflicto, pero anuncia que está dispuesto a llegar a él para lograr un acuerdo favorable. Este hecho lleva a que el receptor del mensaje se cuestione el nivel de compromiso de su oponente, si está diciendo la verdad, o si solamente es un farol. Por lo tanto, el emisor debe estar totalmente comprometido a ejecutar las acciones establecidas en el mensaje, o por lo menos debe hacerle creer al receptor que efectivamente lo hará. Con estas consideraciones, se dirá que una amenaza es creíble si se cumple alguna de las siguientes condiciones: 1) los mensajes son vinculantes, o 2) si los mensajes son no vinculantes, entonces la probabilidad de decir la verdad debe ser alta. La primera es una condición suficiente ya que cuando los mensajes son vinculantes el problema de la credibilidad es irrelevante. La segunda condición es menos restrictiva, y simplemente implica que el receptor debe creer en el significado literal del mensaje si y solo si es altamente probable que le estén diciendo la verdad.

1.2. Sobre Negociación y Comunicación

El problema de negociación del excedente no es nuevo en la literatura económica. Si existe una cantidad de dinero que debe ser dividida entre dos personas, ¿Quién se quedará con la parte más grande del pastel? Schelling (1956) en su *Essay on Bargaining* aborda el problema haciendo una observación general: no es un juego de suma cero. Es claro que los intereses de los jugadores están en conflicto cuando se comparan dos asignaciones eficientes en el sentido de Pareto. Sin embargo, el rompimiento de la negociación siempre es un resultado factible, y si se compara esto con los beneficios de alcanzar cualquier acuerdo, los intereses de los jugadores están en la misma dirección.

La siguiente observación de este autor es que un negociador puede asegurarse un resultado favorable implementando una amenaza. Para ilustrar esta idea, Schelling usa un ejemplo extremo. Cuando dos camiones llenos de dinamita se encuentran frente a frente, en una carretera donde solo uno de ellos puede pasar, ¿Quién debe dar marcha atrás para dejar pasar al oponente? Desde un punto de vista táctico, si uno de ellos toma la iniciativa de avanzar en primer lugar, dejando ver su postura agresiva y una dramática expresión facial de ira, entonces la negociación habrá terminado, el primero en anunciar una amenaza creíble será el ganador. De acuerdo con autores como Frank (1988), Hirshleifer (1987) y Elster (1996, 1998), las emociones pueden ser usadas para ganar credibilidad. Es mejor ejecutar este tipo de jugadas estratégicas con "la sangre caliente".

En esta investigación se tomará como referencia la teoría del compromiso propuesta por Schelling (1960), quien introdujo un enfoque táctico para estudiar la comunicación y la credibilidad en teoría de juegos[1]. A partir de estas ideas pioneras, Hirshleifer (1987, 2000) y Klein y O'flaherty (1993) han trabajado en el análisis y caracterización de las jugadas estratégicas. En la misma dirección, Crawford y Sobel (1982) formalmente mostraron que un agente puede revelar su información privada con el fin de inducir cierto comportamiento en su oponente. Farrell (1993) y Farrell y Rabin (1996), resaltan que el principal problema al modelar la comunicación no vinculante está en el equilibrio *babbling*, donde los mensajes carecen completamente de significado. Estos autores han mostrado que el *cheap talk* puede transmitir información en un contexto general de señalización, encontrando un equilibrio particular en el cual se cumple esta condición. En la misma línea, Rabin (1990) desarrolla el concepto *credible message profile*, buscando un equilibrio particular donde los mensajes tienen un significado concreto y afectan el comportamiento.

En cuanto a la contrastación empírica de estas ideas, la efectividad de la comunicación para alcanzar un acuerdo eficiente ha sido un tópico de gran controversia. Es claro que sin comunicación es difícil alcanzar un resultado eficiente en juegos de coordinación. Cooper, et al (1992) presenta evidencia experimental de la importancia de los mensajes no vinculantes en este tipo de juegos. Wilson and Sell (1997) encuentran un resultado similar en un problema de bienes públicos. Por su parte, Charness (2000) muestra que modificar el orden en que se elige la acción y el mensaje puede tener efectos sobre el comportamiento. Forsythe et al. (1991) encuentran que la comunicación vía *cheap talk* no tiene un efecto significativo en contextos de negociación. Por el contrario, Valley et al. (1998) muestran que el *cheap talk* si puede ser utilizado por los negociadores para llegar a un mejor acuerdo.

El juego del ultimátum y el modelo de Rubinstein son los contextos más frecuentemente utilizados para estudiar la negociación por medio de experimentos económicos. El juego del ultimátum es una representación estilizada de una amenaza. El mecanismo es el siguiente: un emisor comunica una división factible del excedente, dejándole claro al receptor que solo tiene dos opciones: aceptar en las condiciones establecidas, o llegar al punto de desacuerdo si rechaza. Este juego es la base para estudiar la negociación tanto en ambientes de información perfecta (Bolton, 1991; Fehr y Schmidt, 1999; Bolton y Ockenfels, 2000, entre otros) como en situaciones

[1] Para una visión general de la contribución de Schelling a la teoría económica, ver Dixit (2006) y Myerson (2009).

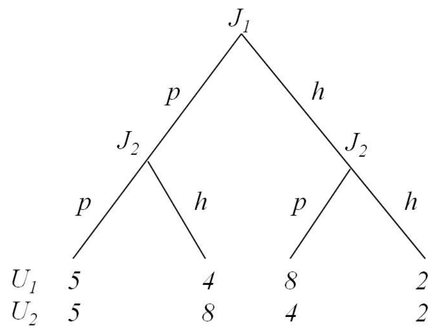
con información imperfecta (Mitzkewitz y Nagel, 19993, Kagel et al., 1996; Straub y Murninghan, 1995; Croson, 1996). Por su parte, el modelo de Rubinstein (1982) es un juego de negociación con ofertas y contraofertas, generalizando la idea del ultimátum. En este sentido, la evidencia presentada por Binmore et al. (1985) ratifica la idea de que las amenazas afectan la división del excedente solamente cuando las divisiones propuestas hacen endógenamente creíble para los jugadores ejecutar la amenaza.

En esta misma línea, algunas hipótesis sobre la efectividad de mensajes estratégicos han sido estudiadas en otros contextos. Güth et al. (1998) y Huck y Muller (2005) estudian el efecto del orden en la posición cuando se negocia de forma secuencial, y las ventajas de quemar dinero en el juego de la batalla de los sexos. Güth et al. (2006) estudian el comportamiento estratégico de líderes y seguidores en un experimento de duopolio con observación imperfecta de los movimientos. Fischer et al. (2006) y Poulsen y Tan (2007) han examinado la transmisión de información en el juego del ultimátum. Más recientemente, Poulsen y Ross (2010) han centrado su análisis en el juego de negociación de Nash y Kimbrough y Sheremeta (2010) estudian la posibilidad de usar pagos laterales para evitar la confrontación.

1.3. El Juego de Halcón y Paloma con Información Perfecta

Para estudiar las variables que explican que un negociador tenga un mayor poder sobre su adversario es necesario simplificar el problema, especificando claramente las reglas del juego y las preferencias de los agentes. Por lo tanto, a lo largo de esta investigación se utilizará el *juego de halcón y paloma con información perfecta*, como aparece en la Figura 1.1. Bajo este esquema se usa la metáfora de la *paloma* para denotar un comportamiento cooperativo, y el *halcón* representará un comportamiento agresivo o egoísta. Este juego modela un ambiente de cooperación y conflicto simultáneamente, debido a que los jugadores comparten un interés común que es evitar caer en el resultado *halcón-halcón*, pero no es claro quién debe ceder para darle la ventaja a su oponente. Decimos que hay un conflicto de intereses porque en equilibrio los pagos no son equitativos, es decir, el jugador que toma una posición agresiva pone los resultados a su favor en detrimento del adversario.

Figura 1.1: Juego de halcón y paloma con información perfecta



La predicción teórica por inducción hacia atrás es que el Jugador 2 elegirá de acuerdo con los incentivos la acción *halcón*, cuando observa que el Jugador 1 elige *paloma*; y *paloma* condicionado a *halcón*. Por su parte, el Jugador 1, anticipando este comportamiento, jugará su mejor respuesta que es *halcón*. Por lo tanto, los pagos finales serán $(8, 4)$, a favor del Jugador 1. Esta combinación de estrategias constituye el único Equilibrio Perfecto en Subjuegos, lo que muestra que en este protocolo secuencial existe una clara ventaja por jugar primero.

Conceptualmente, Schelling (1960) distingue entre dos diferentes tipos de jugadas estratégicas: compromisos simples y amenazas. El compromiso simple implica tomar la ventaja jugando primero, anunciando de forma creíble que la decisión ya está tomada y que es imposible echarse atrás. Por su parte, las amenazas son movimientos de segundo jugador, donde convincentemente se establece una regla de respuesta ante los posibles movimientos del adversario. La característica distintiva de las amenazas es que el emisor del mensaje no tiene incentivos a cumplir con lo establecido, pues por ganar poder de negociación se estaría autoimponiendo algunos costos. Por lo tanto, para lograr que la amenaza sea creíble, es muy importante hacer público que la regla de respuesta no se puede modificar, porque es demasiado costoso, o incluso imposible cambiar de decisión (ver Hirshleifer, 2000).

Muy brevemente podemos resaltar tres mecanismos que sirven para lograr una amenaza creíble, y que se discuten ampliamente en la literatura sobre contratos y negociación. El primer mecanismo consiste en cambiar las ganancias del juego, lo cual generalmente debe ser implementado destruyendo los pagos propios. La segunda posibilidad consiste en utilizar a otros para ayudar a mantener el compromiso. Un equipo puede resultar más fácilmente creíble que un individuo. Por último, se puede cambiar el juego de tal forma que la capacidad de desviarse de un compromiso se vea limitada. La forma más radical de hacerlo es negarse a uno mismo toda oportunidad de cambiar de opinión (ver Dixit y Nalebuff, 1993). Dentro de los límites de esta investigación, este último mecanismo será la base para analizar empíricamente los elementos constitutivos de las amenazas: los mensajes sobre el plan de acción y su credibilidad.

¿Puede el Jugador 2 mejorar su poder de negociación en este contexto? La teoría del compromiso muestra que sí es posible, amenazando de forma creíble con ser un negociador agresivo. Para simplificar el análisis, se supondrá que el Jugador 2 puede enviar mensajes vinculantes sobre su plan de acción[2]. Dado este caso, ¿Qué mensaje debería enviar? Una amenaza consiste en anunciar que se va a responder siempre como *halcón*, independientemente de las acciones del rival. El Jugador 2 amenaza con estar dispuesto a llegar al peor resultado posible, en caso que el Jugador 1 decida tomar ventaja de su posición. Al observar este mensaje, el Jugador 1 preferirá jugar de forma menos agresiva, llevando a los pagos $(4, 8)$, ahora a favor del Jugador 2.

Adicionalmente, existe otra estrategia interesante. El emisor podría estipular que jugará de forma recíproca, cooperando si el otro lo hace, pero dejando claro que el comportamiento agresivo se castiga de igual forma. Este anuncio tiene las características de una amenaza, pero es mejor decir que es una promesa porque establece premiar la cooperación con una acción contraria a los incentivos. Aunque esta estrategia tiene serios problemas de credibilidad, tiene la ventaja de eliminar el conflicto sobre la distribución del excedente, pues el Jugador 1 deberá escoger entre los resultados equitativos $(5, 5)$ o $(2, 2)$.

Con estas consideraciones, ¿Cómo cambiaría la negociación si los mensajes son no vinculantes? En términos generales, esta especificación implica modelar la comunicación como *cheap talk*, es decir, mensajes que no afectan la estructura de los pagos. Desafortunadamente, con base en el Equilibrio Perfecto en Subjuegos, la predicción

[2]Para evitar confusiones, los jugadores se denominan 1 y 2 para denotar el orden en que se toman las decisiones. Adicionalmente, se dirá que el Jugador 2 será el emisor de los mensajes, lo cual implica incorporar una etapa previa al juego.

sería la misma que en el caso de no tener mensajes. Además, el Jugador 2 será indiferente entre los 4 mensajes posibles y, por lo tanto, no hay ninguna predicción específica con respecto a cuál escogería, incluso no está claro si debería enviar el mensaje o no.

Intuitivamente se esperaría que el Jugador 2 envíe aquellos mensajes que siendo vinculantes, mejoran su poder de negociación. Sin embargo, puede mentir estratégicamente, es decir, anunciar una amenaza, pero en el momento de ejecutar las acciones escoger siempre la mejor respuesta. Con los elementos expuestos, es posible resumir estos resultados teóricos en 3 hipótesis contrastables: 1) El poder de negociación del emisor es el mismo cuando no hay mensajes que en caso de contar con mensajes no vinculantes. 2) El poder de negociación del emisor es mayor cuando puede enviar mensajes vinculantes, dado que esto le permite anunciar un compromiso creíble. 3) El emisor mentirá estratégicamente cuando los mensajes son no vinculantes.

1.4. Líneas Generales de la Tesis

La investigación aquí propuesta contribuye a la literatura sobre negociación y comunicación de dos formas complementarias: por un lado, mediante la contrastación empírica y, por otro, reformulando los conceptos teóricos basados en la evidencia. La contrastación se basa en la metodología de la economía experimental, es decir, se recrea una situación de negociación en un ambiente controlado. En este sentido, se pueden identificar tres contribuciones principales.

Primero, para capturar el efecto de las amenazas sobre el poder de negociación del emisor, se propone un novedoso diseño experimental basado en el *juego de halcón y paloma con información perfecta*. Este juego ha sido estudiado experimentalmente por Bornstein et al. (1997), Duffy y Feltovich (2002) y Neugebauer et al. (2008). Sin embargo, el poder de negociación no es central en ninguno de esos trabajos.

Segundo, se ha identificado el efecto particular de los mensajes y la credibilidad sobre la distribución de los pagos. Para llevar esto a cabo, se han definido tres tratamientos experimentales relacionados con el uso estratégico de la comunicación: 1) no hay posibilidad de enviar mensajes, 2) los mensajes son no vinculantes, y 3) los mensajes establecen un compromiso. Adicionalmente, se captura el efecto de la experiencia en la negociación comparando el comportamiento en un primer juego inicial, y si este cambia a medida que la interacción se repite.

Tercero, se hace una contribución a la discusión de los efectos del procedimiento experimental sobre el comportamiento, siguiendo el trabajo de Brandts y Charness (2000). Para ello se definen dos tratamientos: la negociación secuencial (*hot*) y el método de la estrategia (*cold*). La ventaja de esta comparación es que permite capturar endógenamente el efecto emocional en la toma de decisiones estratégicas. De esta forma, el diseño experimental permite analizar el efecto tratamiento *hot-cold* en los diferentes factores anteriormente mencionados: mecanismo de comunicación y experiencia[3].

De forma complementaria, la evidencia empírica encontrada en esta investigación ha permitido identificar algunos vacios conceptuales y ha servido de base para formular nuevas hipótesis de investigación a nivel teórico. Por lo tanto, se contribuye a la literatura teórica de la comunicación estratégica de dos formas. Primero, proponemos una caracterización particular de *advertencias*, *amenazas* y *promesas* como categorías mutuamente excluyentes. Para esto, primero se plantea el *juego del conflicto con información perfecta*, un protocolo de juego secuencial basado en el *juego del conflicto* 2×2 originalmente propuesto por Baliga y Sjöström (2004). Este juego de referencia será útil porque es un modelo estilizado que captura diferentes niveles de alineación en las preferencias, donde juegos clásicos como *cazar ciervo*, *halcón y paloma* y *el dilema del prisionero* son considerados como casos particulares.

Segundo, se modela las jugadas estratégicas con mensajes no vinculantes, mostrando que la decisión de enviar un mensaje en particular y su credibilidad están relacionadas con el nivel de conflicto del juego. En este sentido, el *juego del conflicto con mensajes no vinculantes* captura una situación de negociación donde los agentes pueden hablar sobre sus intenciones de juego, sin que esto tenga ningún efecto directo en los pagos. Se analiza un juego con información imperfecta donde el segundo jugador (el emisor) puede comunicar su plan de acción al primer jugador (el receptor).

[3]En resumen, la metodología aquí propuesta captura el poder de negociación de las amenazas en un diseño experimental factorial $3 \times 2 \times 2$. La primera variable es la comunicación, sus factores son: 1) sin mensajes, 2) mensajes no vinculantes y 3) compromiso. La segunda variable es la experiencia, sus factores son: 1) sin experiencia y 2) experiencia tras 10 repeticiones del juego. Por último, el procedimiento experimental para capturar el comportamiento tiene dos factores: 1) negociación secuencial y 2) método de la estrategia.

Con los elementos anteriormente presentados, los resultados empíricos responden a las siguientes preguntas: 1) ¿Los negociadores hacen uso de la ventaja táctica que implica el compromiso para manipular el comportamiento del rival? 2) ¿La experiencia en la negociación es necesaria para llevar a cabo una amenaza? 3) ¿Los sujetos mienten estratégicamente para poner el acuerdo a su favor? La respuesta a las dos primeras preguntas está directamente ligada. Cuando se compara la negociación con y sin compromiso, encontramos que los sujetos experimentales no identifican las ventajas de usar una amenaza en el primer periodo. Sin embargo, cuando la experiencia aumenta, los sujetos entienden el rol estrátegico de la comunicación, poniendo el acuerdo a su favor.

Por su parte, para responder la tercera pregunta se compara la negociación cuando hay compromiso y cuando los mensajes son no vinculantes. Se encuentra que la credibilidad es esencial para ganar poder de negociación. En general, los mensajes no coinciden con las acciones, lo que implica que las personas mienten de forma estratégica. Es importante aclarar que no se miente de forma aleatoria, sino que existe una clara intención, inducir al oponente a cooperar, pero no se cumple la amenaza cuando esto implica alcanzar un pago menor. Sin embargo, este comportamiento termina siendo perjudicial para el emisor, debido a que cuanto más se miente, menos creíbles son los anuncios.

Con respecto al procedimiento experimental, la evidencia de Brandts y Charness (2000) muestra que la forma de implementar el juego secuencial no tiene un gran impacto en el comportamiento de los sujetos. Este mismo resultado se encuentra cuando se considera la muestra únicamente para la primera interacción. El comportamiento es similar en los dos procedimientos, independientemente de los mensajes o su credibilidad. Por el contrario, cuando los sujetos ganan experiencia, el tratamiento *hot* conlleva un efecto negativo sobre el poder de negociación del emisor, y el peor escenario para él es la negociación secuencial sin mensajes.

Reiterando el resultado anterior sobre la credibilidad, existe una excepción al efecto negativo del tratamiento *hot* cuando se negocia con mensajes no vinculantes. Los sujetos suelen mentir, pero en la negociación estrictamente secuencial se cumple con las acciones especificadas en el mensaje en una mayor proporción, y por lo tanto se tiene un mayor poder de negociación. Este hecho puede ser explicado por los efectos emocionales y porque las mentiras son más difíciles de observar en el protocolo secuencial. Para ganar credibilidad, la evidencia muestra que "amenazar es un plato que se sirve mejor caliente".

Como puede observarse, la evidencia empírica va en la misma dirección que la teoría del compromiso. Sin embargo, los hallazgos sobre el comportamiento han sido de gran utilidad para pensar nuevamente las categorías de análisis. Por lo tanto, se ha estudiado conceptualmente la importancia de tres elementos de la teoría del compromiso a nivel teórico: escoger una regla de respuesta, anunciar estas acciones futuras y la credibilidad de los mensajes. Los resultados teóricos se pueden organizar en tres preguntas centrales: 1) ¿Cuál es la motivación detrás de amenazas y promesas? 2) ¿Pueden los mensajes vinculantes aumentar el poder de negociación del emisor? 3) ¿Cuán creíble debe ser la amenaza pare que la jugada estratégica tenga éxito?

Se han definido amenazas y promesas en el *juego del conflicto secuencial con mensajes no vinculantes*. Esto ha permitido identificar que la motivación para llevar a cabo estas jugadas es alcanzar un pago mayor para el emisor, comparado con el que alcanzaría si la negociación se llevará a cabo sin comunicación. Adicionalmente, dado que el mensaje debe especificar al menos una acción que no sea la mejor respuesta, estas jugadas estratégicas son más relevantes cuando el nivel de conflicto es alto. En general, se demuestra que los mensajes vinculantes aumentan el poder de negociación del emisor en el *juego del conflicto con información perfecta*.

El supuesto de mensajes vinculantes se usa como un paso intermedio para modelar la credibilidad. Este supuesto sirve como línea base para el análisis de los mensajes no vinculantes. Por lo tanto, el siguiente paso es reconocer que la credibilidad está relacionada con que tan probable es que el emisor efectivamente ejecute las acciones especificadas en los mensajes no vinculantes. Para llegar a formular la respuesta a la tercera pregunta, es importante resaltar que los jugadores deben compartir un lenguaje común, donde el significado literal del mensaje puede ser la forma de evaluar su credibilidad. Por lo tanto, el receptor debe creer en el significado literal del mensaje si y solo si es altamente probable que se esté diciendo la verdad.

Técnicamente, la intuición detrás de las amenazas creíbles se captura en dos axiomas: *creencia de que el mensaje es verdad* y *poder de negociación del emisor*. Con estos elementos, se calcula el Equilibrio Bayesiano Perfecto que cumple con estos axiomas, mostrando que los mensajes no vinculantes pueden revelar la información privada cuando el conflicto es bajo. Por el contrario, cuando el conflicto es alto existirán incentivos muy fuertes a mentir, y el *cheap talk* carecerá completamente de significado. Sin embargo, incluso en la peor situación, los mensajes no vinculantes pueden tener un significado preciso si se centra la atención en la posibilidad de cumplir con amenazas o promesas. Como se postula de forma general, las amenazas deben ser creíbles, pero no necesariamente al cien por cien. El modelo propuesto permite establecer con precisión el nivel de credibilidad dependiendo de los incentivos en cada juego. Cuanto más conflicto exista, más credibilidad se debe tener para que la amenaza sea efectiva.

Este trabajo de investigación se organiza en tres artículos, que corresponden a los capítulos de esta tesis. En el capítulo 2 se presenta el diseño experimental basado en el *juego de halcón y paloma con información perfecta*, mostrando los resultados encontrados sobre el poder de negociación de las amenazas en un ambiente *hot*. El capítulo 3 muestra los resultados de un experimento que replica la metodología inicial, pero utilizando el método de la estrategia. Por lo tanto, los efectos tratamiento *hot-cold* se analizan comparando las diferencias en el comportamiento entre los dos experimentos. Por último, el análisis teórico sobre la credibilidad se estudia en el capítulo 4. Este artículo final busca conceptualizar la credibilidad como una variable continua, generando predicciones más precisas en caso de comunicación no vinculante.

Chapter 2

The Bargaining Power of Threats in the Sequential Hawk-dove Game

Abstract: This article studies experimentally the conditions that improve bargaining power using threats in a negotiation process. The analysis is focused on three essential elements of commitment theory: the possibility of a player to announce his own actions, the credibility of these messages, and the experience acquired in the negotiation process. For this aim, we choose the *hawk-dove game with perfect information* because it is a stylized negotiation environment with an unequal distribution of wealth in equilibrium. The experimental data shows that, in the first period, subjects are not aware of the bargaining power of commitment. When the game is repeated and experience increases, subjects quickly understand the advantages of threatening, turning the payoffs into their favor. The credibility of messages is also relevant, given that in some cases subjects lie strategically, reducing their own bargaining power.

JEL Code: C91, D03, D74

Keywords: Credible threats, negotiation, experiments.

2.1. Introduction

Conflict and bargaining are inextricably linked. This can be observed in a wide and varied set of real life interactions. States engage in wars over territories, legal disputes end up in trial, unions go into strikes, firms fight price wars against their competitors, couples argue on the distribution of chores. It is often recognizable that if an agreement cannot be reached, the resulting collision would leave the conflicting parties in a worse scenario. This generates incentives to increase one's bargaining power unilaterally using threats. Consider the case of sending a message, in which one gives the other party the opportunity to choose between a *non-violent* outcome or a fight. This threat is the communication of one's own incentives, designed to determine on the other the acknowledgment of the consequences of his actions, which in case of succeeding in deterring him, will incidentally benefit both parties.

Schelling (1960) distinguishes between two different types of strategic moves: ordinary commitment and threats. The ordinary commitment is the possibility of playing first, announcing that the decision has already been taken, and that it is impossible to be changed, which forces the opponent to make the final choice. On the other hand, threats are second player's moves, where he convincingly pledges to respond to the opponent's earlier choice in a specified contingent way. The distinctive feature of a threat is that the sender has no incentive to carry it out either before or after the event. Therefore, for credibility sake, it is necessary to publicly state that the move cannot be changed, given it is too costly or even impossible to turn back (see Hirshleifer, 2000)[1].

In the experimental literature, Rubinstein's (1982) dynamic model and the ultimatum game are the most popular environments to study bargaining. The evidence found by Binmore et al. (1985) supports the non-cooperative view that outside threats affect bargaining divisions only if divisions proposed make it credible for players to exercise the threat. The ultimatum game is the base for studying bar-

[1] Briefly, we can highlight some mechanisms to reach a credible commitment, which are discussed widely in the literature on contracts and negotiation. The mechanism analyzed here consists of changing the game in such a way that the possibility to cheat could be limited. The most radical situation is to reject any opportunity to go back. Another possibility consists of changing the game payoffs, which can be implemented by rewards or punishments, or by destroying own payments. Finally, it is possible to use other people to support the commitment. A team can be easily more credible than an individual (see Dixit and Nalebuff, 1993).

gaining using perfect information (Bolton, 1991; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000, among others.) although there have been approaches with imperfect information (Mitzkewitz and Nagel, 1993; Kagel et al., 1996; Straub and Murnighan, 1995; Croson, 1996).

In the same way, some hypotheses about the effectiveness of self-serving messages have been tested previously. Güth et al. (1998) and Huck and Muller (2005) studied the positional order effect, and the advantages of burning money in the battle of the sexes game. Güth et al. (2006) studied the strategic behavior of leaders and followers in sequential duopoly experiments with imperfect movement observations. Fischer et al. (2006) and Poulsen and Tan (2007) tested the information transmission in the ultimatum game. More recently, Poulsen and Roos (2010) focused their analysis in a sequential Nash's demand game, and Kimbrough and Sheremeta (2010) studied side-payments for avoiding a lottery contest.

Our article contributes to this literature in two ways. First, in order to capture the effect of threats over the sender's bargaining power, we propose a novel experimental design based on the *hawk-dove game with perfect information*. The convenience of using this game is that it is a stylized representation of preemption and deterrence, and it is the simplest situation for backward induction reasoning[2]. Second, we identify the particular effect of messages and credibility on the final outcome. For this aim, we define three different treatments related with the strategic role of communication: one without communication; one where subjects can announce their intentions, but messages do not reveal necessarily the action chosen; and finally, one where subjects can send messages and commit themselves.

This article focuses on the following questions: *Do people take advantage of commitment opportunity in order to manipulate their opponents' behavior?* and, *is experience in the bargaining process necessary to implement the threat strategy?* By comparing between a negotiation with and without messages, we find that subjects are not aware of the bargaining power of threats in the first period. Nevertheless, when the game is repeated and experience increases, subjects understand the advantages of threats, and turn the payoffs into their favor. For an additional control, we analyze the negotiation with cheap talk messages, and through it we want to

[2]This game was studied experimentally by Bornstein et al. (1997), Duffy and Feltovich (2002) and Neugebauer et al. (2008). However, their focus was not on bargaining power.

answer *whether people would lie strategically to turn the agreement into their favor?* The comparison between the *non-binding* and *commitment messages* shows that credibility is essential for strategic moves. In some cases subjects lie strategically, reducing their own bargaining power.

The article is organized as follows. In Section 2.2 the experimental design is described, highlighting the hypotheses on rational behavior. Section 2.3 presents the main results obtained from the experiment. Finally, Section 2.4 concludes.

2.2. Experimental Design

The experiment was run in November 2010 with economics and management students at the University of Granada, Spain, for a whole sample of 148 subjects[3]. The experiment was fully programmed and conducted using the software z-Tree (Fischbacher, 2007). The use of computers allows the instructions to be completely presented on each individual screen, and a brief questionnaire was realized to ensure their understanding. We conduct a one-shot game, plus a surprising restart with 10 period repetitions. The possibility of playing repeatedly helps us to solve the problem related to the experimental subjects' lack of experience (see Crawford, 1998).

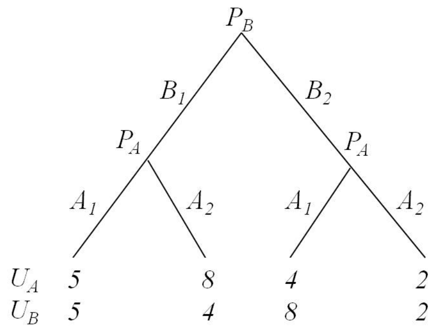
Based on the *hawk-dove game with perfect information*, in Figure 2.1, we recreate a situation where Player A can send a pre-play message, choosing among four different reaction rules, whereas Player B has only two strategies[4]. In order to controlling for the effect of reputation over credibility, we use the absolute strangers protocol. The role of Player A or B is maintained during the whole experiment, but pairs change each period and players never interact with the same partner twice. It is important to clarify that at the end of each period, participants were completely

[3]Financial support from the Spanish Ministry of Education and Science (grant code SEJ2009-11117/ECON) and the Proyecto de Excelencia (Junta de Andalucía, P07-SEJ-3261).

[4]Treatments allow variations in Player A's set of strategies, and in strict terms the normal form of the *hawk-dove game* was never carried out. However, subjects in the experiment always had the 2x2 payoff matrix, as a reference. This was done for simplicity reasons. Considering that players will have a clearer understanding of the instructions and we could avoid to make explicit references to the game sequence. The payoffs that appeared in the matrix were in Euros, instead of in experimental currency units (ECU). See appendix.

informed about decisions and payoffs, as repetitions looked forth to increase player's experience about the game. Subjects were paid the one-shot game payoffs, plus the payoff of one randomly chosen period out of the 10 periods after the restart. Average payments were 10 Euros.

Figure 2.1: Hawk-dove game with perfect information



Our main focus is the possibility of self-committing in a credible way using messages about intentions. Hence, we have three treatments: one without communication; one where subjects can announce their intentions, but messages are cheap talk; and one where subjects can send messages with their final decision. The characteristics of *without messages*, *non-binding* and *commitment* treatments are summarized in the next subsections.

2.2.1. Baseline: Without Messages

In the baseline treatment (WM hereafter) there are two types of players: Player A and Player B. In the first stage, Player B must choose between two actions, B_1 or B_2. In a second stage, Player A observes the opponent's decision, and he must choose between the two actions, A_1 or A_2. Formally, this sequential game is the *hawk-dove game with perfect information*, as appear in Figure 2.1 above[5].

[5]Player B is the first mover, and Player A is the follower. These labels are in this way because in the other treatments Player A has the possibility to send messages as a pre-play move.

2.2.2. Non-binding Messages

Player A has the possibility to send a payoff-irrelevant message in non-binding messages treatment (NB hereafter). Therefore, we incorporate a new stage at the beginning for this aim. In the pre-play stage (Stage 0), Player A must decide whether he wants to send a message to Player B or not. If Player A *"does not want to send a message"*, Player B will receive the announcement *"there is no message from Player A,"* and the incentives would be exactly the same as in WM. In the other case, when A chooses to send a message, his decision consists on choosing between two actions (A_1 or A_2), for every possible choice of Player B. In other words, Player A chooses a response rule like this: *"If Player B chooses $B1$, I choose (A_1 or A_2); and if he chooses $B2$, I choose ($A1$ or $A2$)"*.

In Stage 1, Player B observes the message, and he must choose between two actions, B_1 or B_2. It is clear for both players that the messages will not necessarily coincide with the actions that define the final outcome. The Player A's payoff relevant action will be made in the last stage. In Stage 2, Player A observes the opponent's decision, and he must choose between two actions, A_1 or A_2.

2.2.3. Commitment Treatment

In the commitment treatment (CT hereafter), the action set to obtain the payoffs and the game sequence are the same as in the NB treatment. The only change is the commitment option. If Player A wants to send a message, he can choose whether the message is his *final decision* or not. In general, it is clear for both players that messages will not necessarily coincide with the actions that define the final outcome. When Player A is choosing that the message is his *final decision*, he is rejecting the opportunity to play in the last stage, given that the payoff relevant actions are now clearly specified in the message. In this case, Player B faces the response rule message, plus a sentence: *"it is his final decision"*.

In summary, Player A chooses whether he wants to send a message or not in Stage 0. If he wants to send a message, Player A choose a particular response rule and whether he wants to bind-himself or not. The possibility of Player A to send a *final decision message* is essential for testing if pre-play communication is useful to increase the sender's bargaining power. In this sense, Player A now has the possibility to play first, committing himself with a reaction rule, and Player B knows that.

The characteristics of WM, NB, and CT treatments are summarized in Table 2.1. Note that Player B's set of strategies and the set of final outcomes are constant among treatments. The changes between treatments are Player A's set of strategies, including the possibility of sending a message, and Player B's information sets.

Table 2.1: Timing in the different treatments

	WM	NB	CT
Pre-play moves			
Stage 0		Player A Message	Player A Message Final decision
Hawk-dove game with perfect information			
Stage 1	Player B	Player B	Player B
Stage 2	Player A	Player A	If no final decision Player A
Theoretical Prediction	(A_2A_1, B_2) $(4, 8)$	(A_2A_1, B_2) $(4, 8)$	(A_2A_2, B_1) $(8, 4)$
Sample	30 subjects	60 subjects	58 subjects

2.2.4. Theoretical Considerations and Hypothesis

In WM and NB the theoretical prediction is that Player A chooses, according to his incentives, strategy A_2A_1. Strictly speeching, Player A chooses A_2 conditioned to B_1, and A_1 conditioned to B_2. Player B plays B_2 given that it is his best response if he forecasts the opponent's behavior, reaching the outcome $(4, 8)$. Player B has more bargaining power given the fact that he has the first mover advantage. We label the message A_2A_1 as *warning*, using Schelling's classification, because that is Player A's strategy in the Subgame Perfect Nash Equilibria.

The theoretical prediction in CT is that Player A is going to play the message A_2A_2, and he is going to reject the opportunity to play in the last stage. We label the message A_2A_2 as a *threat* based on an aggressive statement that intends to deter the opponent of behaving aggressively, too. Faced with this binding threat, Player B's best response is choosing strategy B_1 in order to stay away from the outcome $(2,2)$. In this treatment, we find an essential result in relation to commitment theory, where Player A changes the unfavorable situation to reach the outcome $(8,4)$, which is the highest payoff available for him. Player A's bargaining power is reached by threatening his opponent with being a hostile negotiator, taking the risk of falling into the disagreement outcome.

There is another interesting strategy, the *promise* A_1A_2. Player A might stipulate that he will play reciprocally, cooperating if the other one does it, but clarifying that the aggressive behavior will be punished in an equal way. This message has the characteristics of a threat, but we prefer the label *promise* because it establishes to reward the cooperation with a not best response action. Though this strategy leads to serious problems of credibility, it has the advantage of eliminating the conflict on the surplus distribution. Player B must choose between the equitable results $(5,5)$ or $(2,2)$.

In summary, given that the payoffs matrix is fixed, any variation on the outcome reached between treatments is explained by the availability of messages and their credibility. Based on the commitment theory, Player A has more bargaining power if he is able to threaten his opponent with an aggressive arrangement, and makes it clear that this is the unique relevant option. In consequence, we can expect Player A to use the option of sending the *threat* message because it is the way of turning the outcome into his favor.

Following this argument, we should observe that Player A's largest payoffs are reached in CT, given the availability of binding messages. For this reason, we can compare Player A's average payoffs, and formulate the following hypotheses which captures Player A's bargaining power. For the question:

Do people take advantage of commitment opportunity in order to manipulate their opponents' behavior?

Hypothesis 2.1 *Player A's payoffs are equal in WM and NB treatments.*

Hypothesis 2.2 *Player A's payoffs are larger in CT than in NB treatment.*

In the game we are using for this experiment, total payoffs in equilibrium are 12€ independently of the treatment. Therefore, the average total payoff ($U_A + U_B$) shows the importance of messages to solve the coordination problem, because for both players it is rational to avoid the $(2, 2)$ outcome. This variable captures the efficiency of the three different negotiation procedures, and we expect no significant differences between treatments. For the question:

Is the outcome efficient under the different communication mechanisms?

Hypothesis 2.3 *The average total payoff is not different between treatments.*

As previously stated, the *threat* has to be public and irrevocable before the opponent plays, in order to change the rival's beliefs and his decision. In NB treatment, since Player A can only send non-binding messages, these ones can be used for bluffing. So, although the theoretical prediction is that Player A will choose the action *warning*, he can choose a different message in order to induce the opponent to choose B_1. This is the purpose of threatening. For the question:

Would people lie strategically to turn the agreement into their favor?

Hypothesis 2.4 *In NB treatment, Player A chooses the response rule warning, but he sends the threat message.*

Hypothesis 2.4 must be tested in the direction of Croson et al.'s (2003) results about the possibility of lying and threatening in the ultimatum game. Their conclusion is that non-binding messages have a clear effect on behavior in two different ways. First, cheap talk can influence the counterpart's decision. This result suggests that negotiators are less crafty than the models suggest, but it can be simply related to subjects' inexperience. Second, deceptive messages have a negative impact on outcomes. This suggests that negotiators respond to cheap talk by changing their behavior and they are ready to punish when they feel being cheated, even when such punishment is costly for them.

The last consideration is related to the subjects experience in the bargaining process. We assume that players are utility maximizers and their utility function is defined only over their own payoffs. By modeling commitment, this assumption is important because we can highlight the Subgame Perfect Nash Equilibrium for every treatment. A limitation is that real people usually are not able to do the reasoning associated with backward induction, or at least not as easy as the theoretical model postulate (see Crawford, 1998). Based in the usual assumptions of the rational choice theory, we expect behavior not to be time-dependent. For the question:

Is experience in the bargaining process necessary to implement the threat strategy?

Hypothesis 2.5 *The final outcome achieved does not change across periods.*

2.3. Results

This experiment seeks to stand out if messages, credibility, and experience are essential to improve Player A's bargaining power. To achieve this aim, in Section 2.3.1 we compare the outcome distribution in the one-shot game (OSG) with the outcomes in the last repetition. Furthermore, we analyze the outcome progression over periods in Section 2.3.2. Finally, in Section 2.3.3 we run a logit panel model to understand the variables that influence Player B to cooperate.

2.3.1. One-Shot Game vs. Last Repetition

Crawford (1998) says that behavior in the laboratory often takes time to stabilize; then theories that assume equilibrium are commonly tested by comparing their predictions to observed behavior in the last period. To control for subjects' unfamiliarity with the environment, we provide enough experience via 10-period repetition to assure meaningful responses and to reveal the effects, if any, of learning. Therefore, first we compare the outcomes in the OSG with the last repetition.

Table 2.2: Payoffs distribution

	One-shot Game			Last Repetition (Period 10)		
Outcome	WM	NB	CT	WM	NB	CT
(2,2)	27	10	17	7	17	7
(4,8)	47	53	45	87	53	17
(5,5)	7	10	7	0	3	28
(8,4)	20	27	31	7	27	48
Average	(4.3,5.4)	(5.0,6.0)	(5.0,5.5)	(4.1,7.3)	(4.8,5.8)	(6.1,4.8)

Payoffs distribution are in percentage. Average values are in Euros.

The average player A's payoffs in the OSG were 4.3 € in WM, 5.0 € in NB, and 5.0 € in CT, as shown in Table 2.2. It is clear that payments reached in CT are not significantly different with respect to other treatments (WM and NB, $p = 0.255$; NB and CT, $p = 0.876$)[6]. With regard to efficiency, the average total payoff is 9.7 € in WM, 11.0 € in NB, and 10.5 € in CT. The highest payoff obtained in the *non-binding* treatment is not significantly different with respect to the other treatments (WM and NB, $p = 0.268$; NB and CT, $p = 0.637$).

The evidence in the last repetition supports that Player A's payoffs are equal in WM and NB treatments, and higher in CT. Average Player A's payoffs in Period 10 were 4.1 € in WM, 4.8 € in NB, and 6.1 € in CT. Differences of CT with respect to others treatments are significant at 1% (WM and NB, $p = 0.435$; NB and CT, $p < 0.01$). With regard to efficiency, the average total payoff is 11.5 € in WM, 10.6 € in NB, and 10.9 € in CT. The difference among treatments are not significant (WM and NB, $p = 0.258$; NB and CT, $p = 0.392$).

In the OSG, we reject Hypothesis 2.2 because differences among treatments are not significant. Our conclusion is that played strategies and outcomes depend neither on the possibility of sending messages nor on credibility, if the interaction is carried out only once. However, in the last repetition the picture is completely different. Clearly, the theoretical prediction is very accurate in WM, with the outcome $(4, 8)$ being reached by 87%. In NB, this prediction is reached by 53%. In CT Player A has more bargaining power than in the others treatments, given that $(8, 4)$ is reached by 48% of cases. In the last period, evidence supports Hypotheses 2.1, 2.2 and 2.3. This findings are summarized in Result 2.1

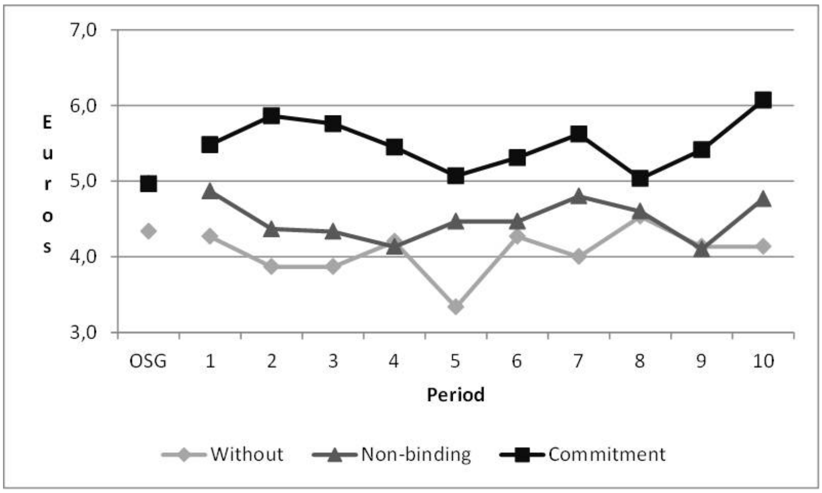
Result 2.1 *Player A does not have more bargaining power in CT if the game is played only once. In contrast, Player A's average payoffs are higher in CT after 10 repetitions. The evidence suggests that some experience is needed in order to use threats in a successful way.*

[6]Mann-Whitney test, two-sided. Between subjects comparisons.

2.3.2. Repeated Game

When the 10 periods are analyzed, Player A's average payoffs are of 4.1 € in WM and 4.5 € in NB, and it is possible to affirm that the trends are completely flat. In Figure 2.2 we see that Player A obtains in NB a higher average payoff than in WM. In the aggregate, the difference is significant at 10% ($p = 0.068$)[7]. In CT this result is stronger, the average Player A's payoffs is always higher than the other two treatments. The average payment is 5.5 €, and this difference between treatments is significant at 1% (NB and CT, $p < 0.001$).

Figure 2.2: Player A's average payoffs



For Periods 1 to 10, the average of the total payoffs is 10.6 € in WM, 9.8 € in NB, and 10.0 € in CT. As in the previous result, it is possible to affirm that the trends are completely flat. However, the efficiency is significantly high in the WM, in contrast to Hypothesis 2.3 (WM and NB, $p < 0.01$; NB and CT, $p = 0.718$).

Result 2.2 *Commitment improves Player A's bargaining power. In the aggregate, the evidence shows the logical treatments order, in terms of Player A's payoffs. In addition, total payoffs are statistically equal in the CT and NB, but in the WM this variable is significantly high.*

On one hand, it is clear that Player A increases his bargaining power in CT, as expected in Hypothesis 2.2. On the other hand, the best condition for Player B

[7] Mann-Whitney test, two-sided. In this section we take the average for the 10 periods repetitions. Between subjects comparisons.

is the negotiation without messages, because the path is going close to the $(4, 8)$ outcome. The question is how does Player A reaches a higher share of the pie? The answer goes in the direction of the theoretical predictions, threatening in a credible way to fall into the worst result if his rival does not cooperate.

Although we are going to study subjects behavior in more detail in the next subsection, descriptive statistics in Table 2.3 provide an overview for treatments comparisons. The first column (B_1) displays the proportion of B players choosing to cooperate in WM. Columns 2 and 7 (labeled with M) correspond to the message distribution in NB and CT, respectively. Regarding NB, Column 3 ($B_1|M$) shows the conditional proportion of B players choosing B_1, given the observed message. Column 4 (*Lie*) is the proportion of A players that choose an action different than the one in the message. Column 5 ($Lie \bigcup B_1$) shows the join proportion of A players that lie after observing Player B chose B_1. Column 6 ($Lie \bigcup B_2$) shows the join proportion of A players that lie after observing Player B chose B_2. Regarding CT, Column 7 ($FD|$M) corresponds to the conditional proportion of A players choosing *final decision*, given the chosen message. Finally, columns 9 and 10 ($B_1|$FD and $B_1|$NFD) are the conditional proportion of B players choosing B_1, given that the observed message is *final decision* and *no-final decision*, respectively.

Table 2.3: Message effectiveness, by treatments

	WM	NB					CT			
	B_1	M	B_1|M	*Lie*	$Lie \bigcup B_1$	$Lie \bigcup B_2$	M	FD|M	B_1|FD	B_1|NFD
A_1A_1		5	7	79	0	79	2	33	0	50
Promise		46	44	36	27	9	36	60	94	57
Warning		6	6	56	0	56	8	39	0	21
Threat		23	32	13	1	12	46	83	74	48
NoMess	11	20	20				8			17
Total	11	100	32	33	16	17	100	63	76	48

Average from Period 1 to 10. Values are in percentage.

A general conclusion from the above table is that A players choose in a higher proportion those messages that give them a higher bargaining power (*threat* and *promise*). Moreover, A Players prefer to send a message (80% in NB and 92% in CT) because not sending any has a low effectiveness, as it is shown by the low percentage of B_1 responses given no messages.

For NB treatment sample, senders are choosing *threat* and *promise* (23% and 46%, respectively) to induce Player B to play B_1. Nevertheless, the effectiveness of those messages is undermined by the fact that some of them behave as strategic liars.

When sending a *threat*, 12% of the subjects lie after observing B_2, and when sending a *promise* 27% of the subjects lie after observing B_1.

In CT the effectiveness of messages is higher than in NB. 46% and 36% of subjects are playing the *threat* and *promise* messages, respectively (83% and 60% of them chooses FD). Beyond doubt, the message in CT is the way to inform about the *threat*, and Player B must choose between the outcomes $(8, 4)$ or $(2, 2)$. Facing this commitment message, 74% of subjects behave like a dove. For *promise* the analysis is similar. Although the effectiveness of *promise* is the highest (94%), this strategy is less profitable given that Player B must choose between the outcomes $(5, 5)$ or $(2, 2)$.

In summary, there is evidence of strategic lying behavior. Senders are choosing those messages that induce Player B to play B_1, but some of them choose the best response action in each case. Clearly, there are strategic liars, but B subjects are not naive believers. The strategic move intends to influence the other player's decision, but without credibility the bargaining power is undermined.

2.3.3. The Strategic Role of Communication

For a deeper analysis, we study Player B's decision as an indicator of Player A's bargaining power. Table 2.4 presents the results of 5 panel logit regressions of Player B's decision to play B_1, including dummy variables for *message*, *final decision*, CT and NB, a linear trend (*Period*), and the time that Player B takes to make his decision (*choosing time*) as regressors. Regressions 2 and 3 include a dummy for the OSG, whereas Regressions 4 and 5 include dummies of Player B's payoff in $t-1$, and the dummy variable Lie_{t-1}, for capturing the fact that Player B could face a lie in the previous period. Regressions 3 and 5 also include the interaction between *period* and each one of the treatment dummies[8].

The possibility of sending messages is a persuasive strong mechanism. As previously stated, the message is used in 80% in NB, and in 92% in CT, on average for period

[8]We also estimated logit regressions clustering by individual to compute the correct statistical significance for these interaction terms (Norton et al., 2004). Notwithstanding, the significance of these interaction terms remains under this specification.

Table 2.4: Regression analysis explaining Player B's decision

Player B (B_1)	(1)	(2)	(3)	(4)	(5)
Message	1.344***	1.396***	1.397***	1.407***	1.407***
	(0.349)	(0.354)	(0.354)	(0.377)	(0.378)
Final decision	1.427***	1.439***	1.371***	1.721***	1.774***
	(0.292)	(0.299)	(0.306)	(0.334)	(0.346)
OSG		-0.428	-0.424		
		(0.359)	(0.355)		
CT	0.497	0.464	-0.005	0.641	0.969
	(0.574)	(0.577)	(0.756)	(0.658)	(0.952)
NB	0.056	0.027	-0.202	0.438	0.518
	(0.551)	(0.554)	(0.749)	(0.626)	(0.937)
Period		-0.029	-0.085	-0.026	0.003
		(0.034)	(0.084)	(0.036)	(0.099)
Period·CT			0.088		-0.056
			(0.093)		(0.113)
Period·NB			0.042		-0.012
			(0.092)		(0.112)
$\text{Payoff}_{t-1} = 8$				0.124	0.113
				(0.291)	(0.293)
$\text{Payoff}_{t-1} = 5$				0.222	0.229
				(0.363)	(0.364)
$\text{Payoff}_{t-1} = 4$				0.065	0.064
				(0.324)	(0.325)
Lie_{t-1}				-0.742**	-0.757**
				(0.370)	(0.373)
Choosing time	0.012**	0.011**	0.011**	0.014**	0.014**
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
N	814	814	814	740	740

Logit panel data estimation with random effects. Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

1 to 10 (the difference is significant at 1%; $p < 0.001$)[9]. As shown in Table 2.4, the *message* coefficient is always positive. In addition, it is clear that the higher Player A's payoffs in CT is explaining for the possibility to send binding messages. The *final decision* variable is always positive and significant. For the first 5 periods, the subjects send the *final decision* message at 53%, but this proportion grows up to 73% for the last five periods.

In order to go one step ahead in this analysis, it is necessary to understand Player B's decision, conditioned to facing a message. In Table 2.5, we perform a regression

[9]Given that sent messages is a binary variable, we use the non-parametrical test of proportions, two-sided.

analysis only for those B Players who faced the message, including dummies of the message sent by Player A as regressors.

Table 2.5: Regression analysis explaining Player B's decision, conditional to $message = 1$

Player B (B_1)	(1)	(2)	(3)	(4)	(5)
Promise	2.761***	3.039***	3.026***	2.932***	2.922***
	(0.672)	(0.705)	(0.702)	(0.807)	(0.807)
Warning	-1.172	-1.207	-1.200	-1.744*	-1.801*
	(0.815)	(0.830)	(0.826)	(1.018)	(1.024)
Threat	1.794***	2.010***	1.992***	1.616**	1.605**
	(0.672)	(0.703)	(0.700)	(0.801)	(0.800)
Final decision	1.266***	1.486***	1.431***	1.875***	1.953***
	(0.358)	(0.374)	(0.381)	(0.420)	(0.433)
OSG		0.031	0.042		
		(0.508)	(0.506)		
CT	1.025**	0.949*	0.650	0.600	1.075
	(0.493)	(0.505)	(0.659)	(0.573)	(0.790)
Period		-0.107**	-0.133**	-0.102**	-0.063
		(0.045)	(0.058)	(0.049)	(0.066)
Period·CT			0.053		-0.080
			(0.075)		(0.091)
$\text{Payoff}_{t-1} = 8$				0.113	0.116
				(0.382)	(0.382)
$\text{Payoff}_{t-1} = 5$				0.375	0.409
				(0.453)	(0.454)
$\text{Payoff}_{t-1} = 4$				0.364	0.380
				(0.415)	(0.416)
Lie_{t-1}				-1.013**	-1.055**
				(0.444)	(0.447)
Choosing time	0.011*	0.006	0.007	0.011	0.010
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
N	557	557	557	505	505

Logit panel data estimation with random effects. Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Player A learns the importance of the *threat* in CT quickly. In Table 2.5 we see that its coefficient is positive and significant. In addition, this message is not the only one to increase Player A's bargaining power. The message *promise* is also used to induce Player's B decision and its effect is also positive and significant. This result is particularly interesting because the strategy *promise* is dominated by *warning*, but can be considered a safety way for obtaining a satisfactory outcome of $(5, 5)$ avoiding the conflict of distribution.

Result 2.3 *The reception of threats as well as promises induces changes on subjects' behavior. Controlling for other variables, the probability of cooperation is larger for those B players who face a message. Besides, this probability is even higher if that message is a final decision.*

In general, Tables 2.4 and 2.5 confirm the previous results, but there are some behavioral variables that we cannot consider in our hypotheses: Lie_{t-1} and *choosing time*. If Player B faces a lie in a previous period, then the probability of cooperate in the current period is reduced. This coefficient is always negative and significant. Lies destroy the messages credibility and lead outcomes to be closer to those that the rational choice theory predicts.

Result 2.4 *Senders are choosing those messages to induce Player B to cooperate, but some of them do not fulfill the agreement. However, deceptive messages have a negative impact on Player A's bargaining power.*

On the other hand, *choosing time* refers to the seconds spent by Player B in order to take the decision. This coefficient is always positive and in Table 2.4 it is always significant. This fact suggest that choosing B_1 is more cognitive demanding for Player B, because it is necessary to spend more time evaluating the consequences. As was finding in Neugebauer et al. (2008), the sequential *hawk-dove game* is a good environment for selfish behavior. Subjects are involving in a bargaining environment, where there are incentives to fight for the higher share of the pie.

With respect to experience, the interaction between *period* and NB or CT variables is not significant in Table 2.4. This fact confirm the flat path of our experimental variables, therefore we cannot reject Hypothesis 2.5 if we consider the whole sample. In Table 2.5 the *period* is negative and significant. As experience increases, the probability of playing B_1 decreases. In summary, subjects are not aware of their bargaining power in the OSG. After that, A players realize the purpose of messages, and turn the outcome into their favor. In addition, B players are not naive believers of messages, as experience increases, they gradually play in a more aggressive way.

Result 2.5 *Subjects A learn how to play after the first period, and then they continue playing in the same way. In addition, when B players face a message, they recognize the opponent's power, but gradually react by playing in a more aggressive* *way.*

2.4. Conclusions

The aim of this article is that of strategic communication in a bilateral bargaining. In accordance with our experimental design, we define three different treatments: *without messages*, *non-binding* and *commitment*. Each treatment clearly separates the particular effect of messages and credibility on the final outcome. In addition, we run 10 repetitions to capture the learning process and the effect where subjects acquire experience in this environment.

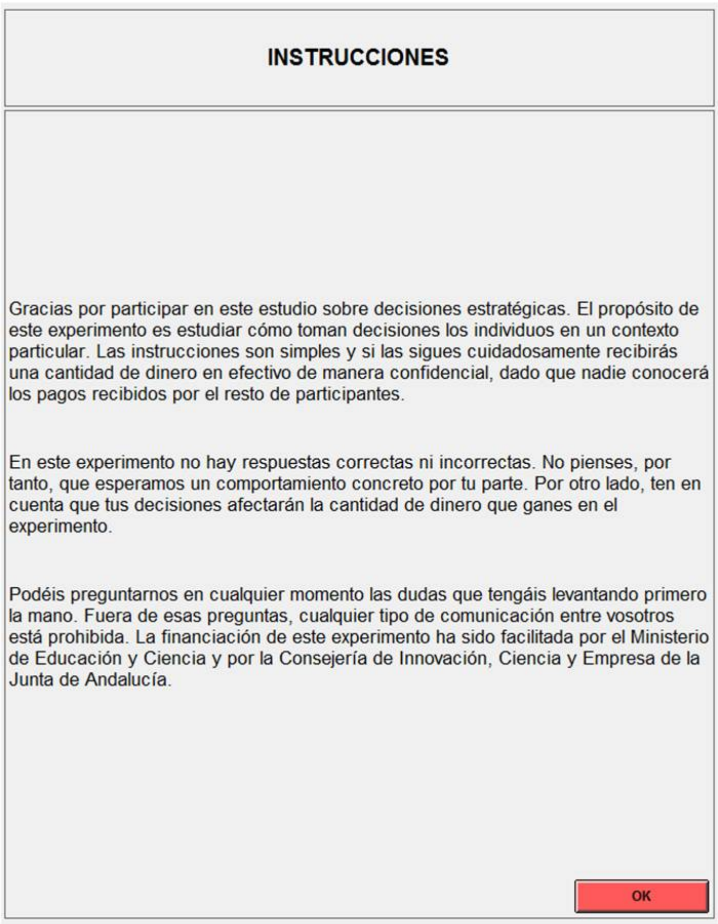
The experimental evidence shows that the use of messages containing intentions improve the sender's bargaining power. Subjects understand the advantages of threatening, by committing in advance in order to influence the opponent's decision. They choose those messages that influence the other to cooperate, even with cheep talk announcements. However, when subjects lie strategically, they loose some of their power. Furthermore, the situation is even worse when they cannot send messages.

Our results are in line with the basic ideas of commitment theory. The simplest strategic move is playing first, and threats are more complex techniques, because the second mover is trying to change his unfavorable position. Subjects acquisition of experience in the bargaining process has a small effect in this sequential game. Subjects learn how to play after the first period and then they continue playing in the same way.

The center of the analysis has been the rational behavior under different tactical alternatives. As expected, the *hawk-dove game* is a good environment for selfish behavior, even more if we take under consideration the usage of the absolute strangers protocol. We find that messages credibility is essential for strategic moves, then, for future research is important to study commitment with endogenous mechanisms of credibility. We must look to implement in the laboratory different ways for obtaining reputation. In the same direction, it is also interesting to extend this procedure to study other payoff environments, as prisoner's dilemma or stag hunt games were mutual cooperation could be implemented in equilibrium.

Appendix: Instructions in the Hot Treatment

Figure 2.3: Screen 1: Common instructions

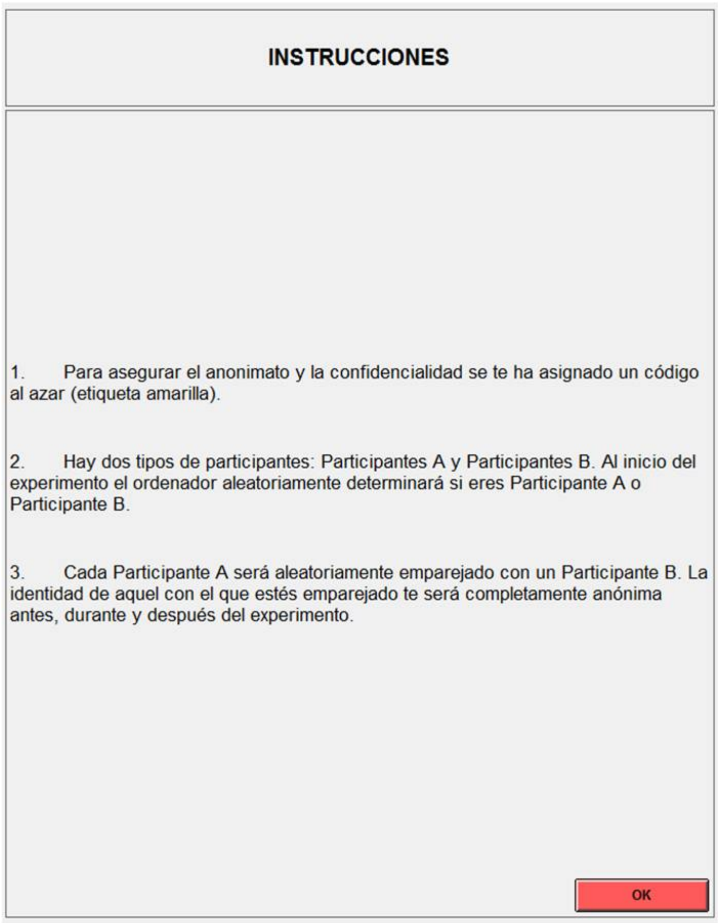


Thank you for taking part in this study on strategic decisions. The intention of this experiment is to study how the individuals make decisions in a particular context. The instructions are simple and if you follow them carefully you will receive an amount of money in a confidential way, so that nobody will know the payments that the other participants will receive.

In this experiment there are nether correct nor incorrect answers. Do not think, therefore, that we expect a concrete behavior for your part. On the other hand, bear in mind that your decisions will affect the quantity of money that you earn in the experiment.

You can ask us at any time about the doubts that you have by raising your hand. Except of these questions, any type of communication among you is prohibited. The financing of this experiment has been facilitated by Ministerio de Educación y Ciencia and by the Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía.

Figure 2.4: Screen 2: Common instructions

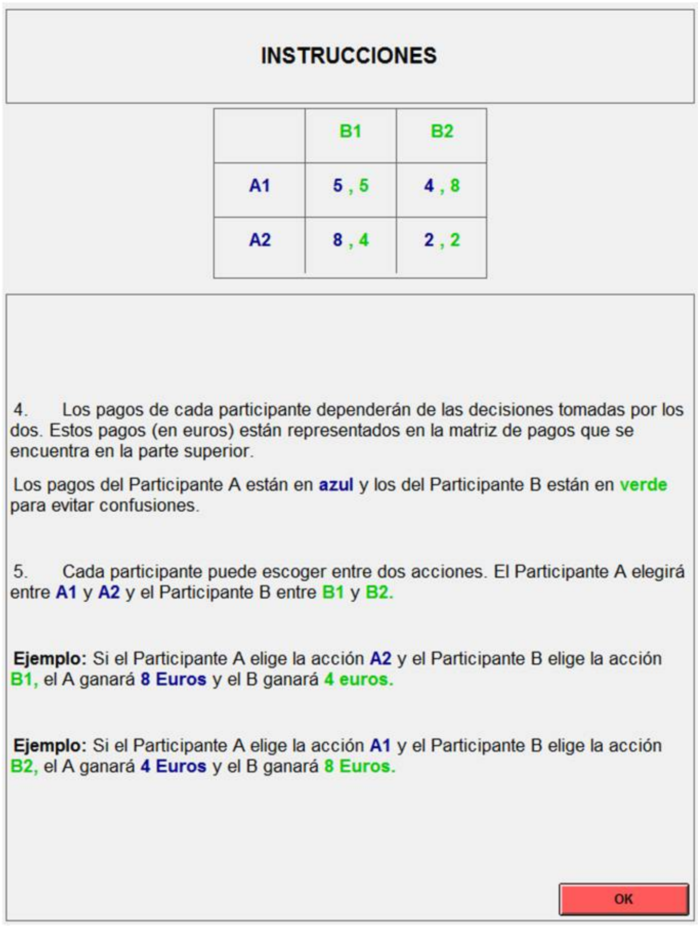


1. To assure the anonymity and the confidentiality a code has been randomly assigned to you (yellow card).

2. There are two types of participants: Participants A and Participants B. From the beginning of the experiment the computer will randomly determine if you are Participant A or Participant B.

3. Every Participant A will be randomly paired with a Participant B. The identity of the participant that is paired with you will be completely anonymous before, during and after the experiment.

Figure 2.5: Screen 3: Common instructions



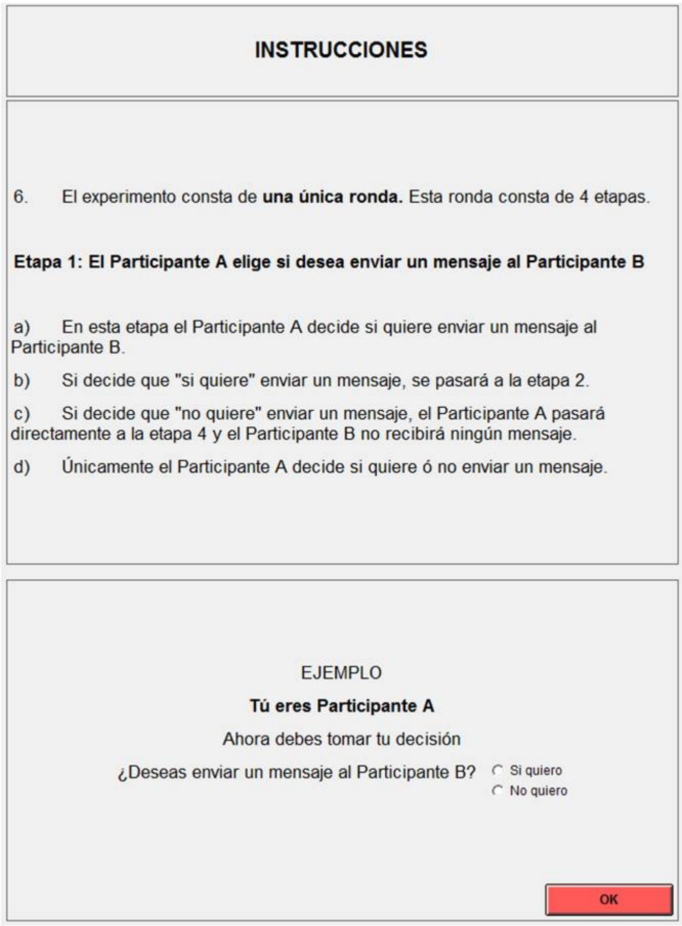
4. The payments of every participant will depend on the decisions made by both. These payments (in euros) are represented in the payoffs matrix that you can find in the top.

5. Every participant can choose among two actions. Participant A will choose among A1 and A2 and Participant B among B1 and B2.

Example: If Participant A chooses the action A2 and Participant B chooses the action B1, A will earn 8 Euros and the B will earn 4 Euros.

Example: If Participant A chooses the action A1 and Participant B chooses the action B2, A will earn 4 Euros and B will earn 8 Euros.

Figure 2.6: Screen 4: CT



6. The experiment consists of only one round. This round has 4 stages.

Stage 1: Participant A chooses if he wants to send a message to Participant B.

a) In this stage Participant A decides if he wants to send a message to Participant B.

b) If he decides that "he wants" to send a message, he will pass to stage 2.

c) If he decides that "he does not want" to send a message, Participant A will pass directly to stage 4, and Participant B will not receive any message.

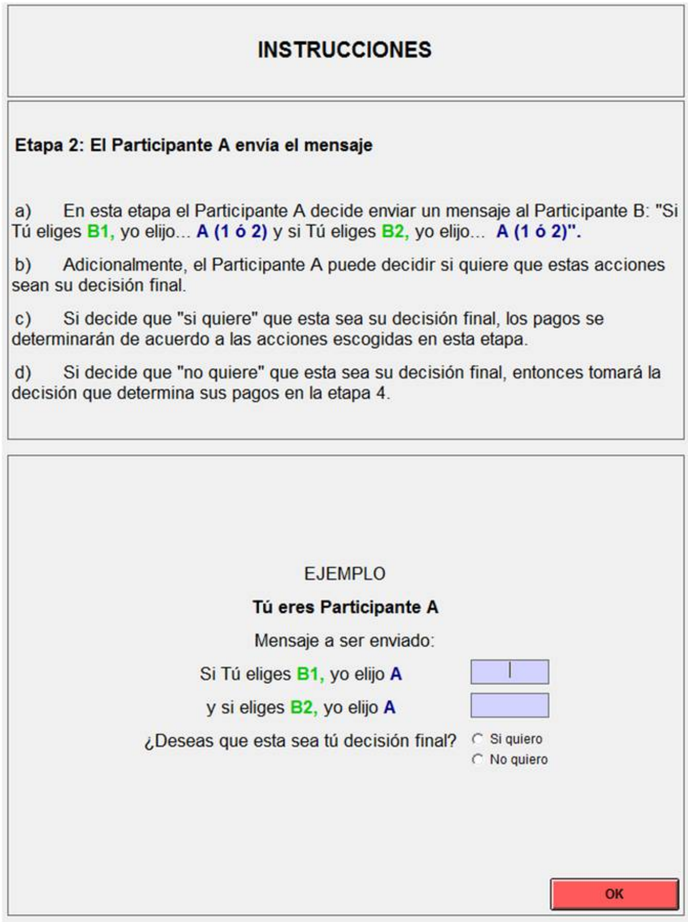
d) Only Participant A decides if he wants to send a message.

EXAMPLE

You are Participant A. Now you must take your decision

Do you want to send a message to Participant B?

Figure 2.7: Screen 5: CT



Stage 2: Participant A sends the message

a) In this stage Participant A decides to send a message to Participant B: “If You choose B1, I choose... A (1 or 2) and if You choose B2, I choose... A (1 or 2)”

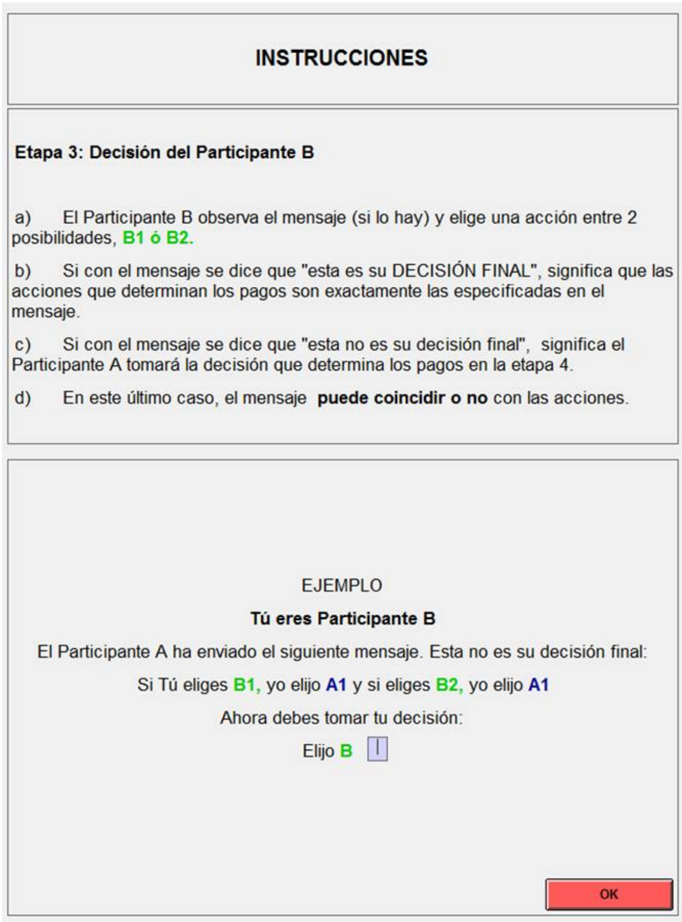
b) Additionally, Participant A can decide if he wants these actions to be his final decision.

c) If he decides that “he wants” that this message is his final decision, the payoffs will be calculated in agreement to the actions chosen in this stage.

d) If he decides that “he does not want” that this one is his final decision, he will make the decision that determines his payoffs in stage 4.

EXAMPLE. You are Participant A. Message to being sent:
If You choose B1, I choose A ( )
If You choose B2, I choose A ( )
Do you want that this message become your final decision?

Figure 2.8: Screen 6: CT



Stage 3: Participant B's decision

a) Participant B observes the message (if any) and he chooses an action among 2 alternatives, B1 or B2.

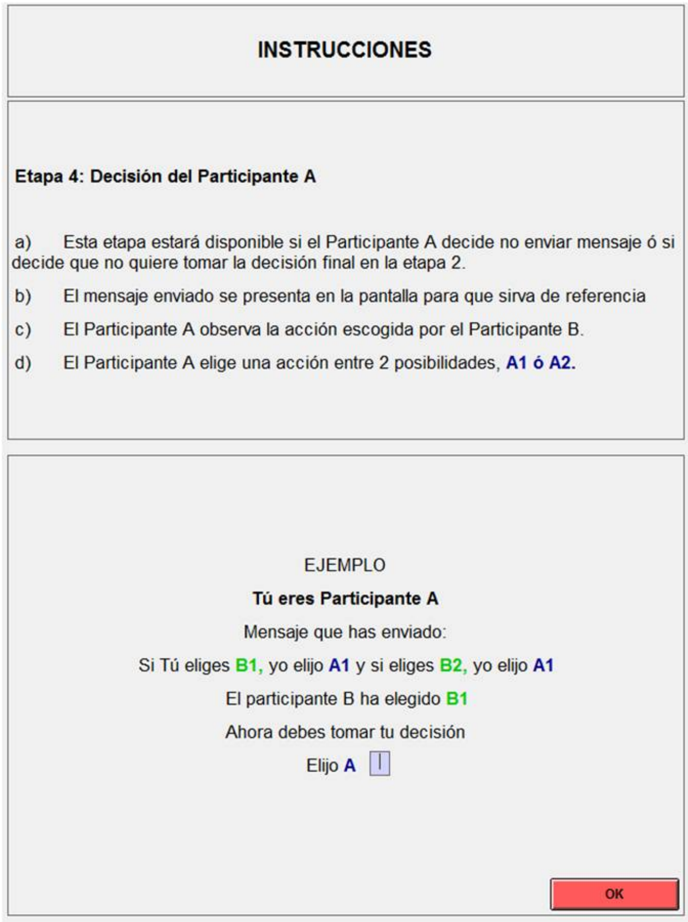
b) If in this message he says that "it is his FINAL DECISION", it means that exactly those actions determine the payments that are mentioned in the message.

c) If in this message he says that "it is not his final decision", it means the Participant A will take the decision that determines the payments in the stage 4.

d) In the latter case, the message could coincide or not with the actions.

EXAMPLE.
You are Participant B.
Participant A has sent the following message. This is not his final decision:
"If You choose B1, I choose A1 and if You choose B2, I choose A1".
Now you must take your decision
I choose B ( )

Figure 2.9: Screen 7: CT



Stage 4: Participant A's decision

a) This stage will be available if Participant A decides not to send message or if he decides not to take the final decision in the stage 2.

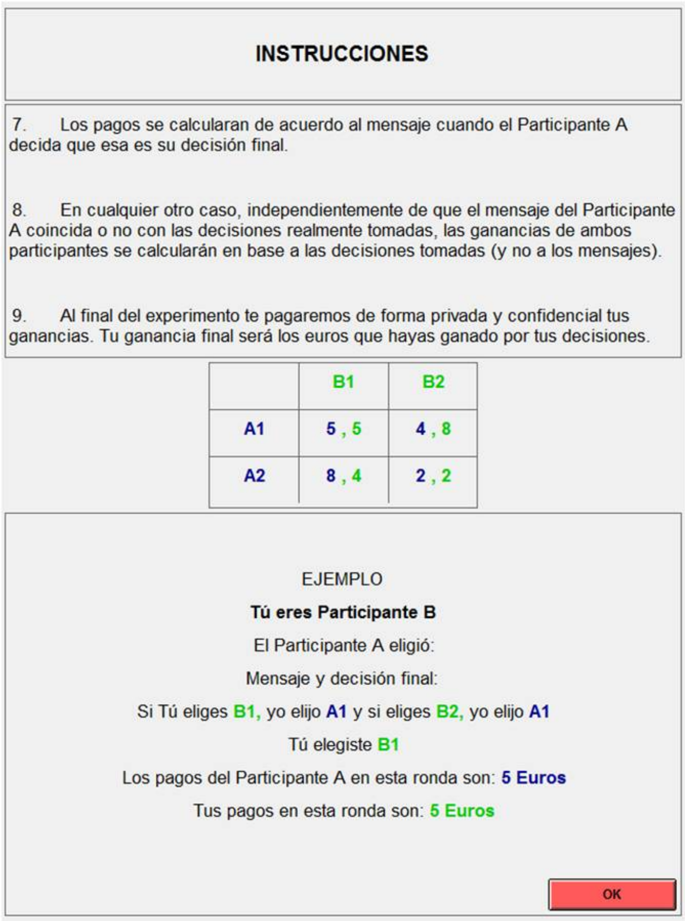
b) The sent message appears on the screen in order to be used as a reference.

c) Participant A observes the action chosen by the Participant B.

d) Participant A chooses an action among 2 possibilities, A1 or A2.

EXAMPLE. You are Participant A.
Message that you have sent: "If You choose B1, I choose A1 and if You choose B2, I choose A1".
The participant B has chosen B1
Now you must take your decision
I choose A ( )

Figure 2.10: Screen 8: CT



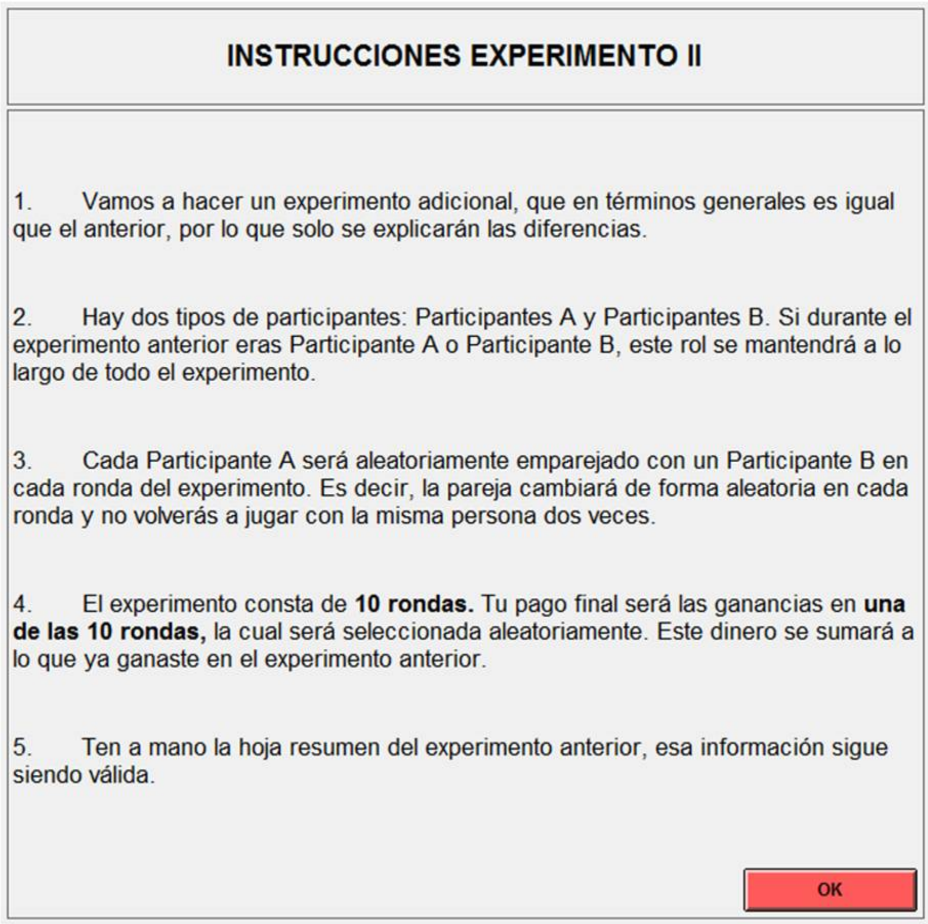
7. Payments are calculated according to the message when the Participant A decides that this it is his final decision.

8. In any other case, independently of Participant A's message coincides or not with the really taken decisions, the earnings of both participants will be calculated based on decisions (not on messages).

9. At the end of experiment we will pay to you in a private and confidential way. Your final payoff will be the euros that you have earned according to your decisions.

EXAMPLE. You are Participant B.
Participant A chose: Message and final decision:
"If You choose B1, I choose A1 and if You choose B2, I choose A1".
You chose B1. Participant A's payoffs in this round are: 5 Euros.
Your payoffs in this round are: 5 Euros.

Figure 2.11: Surprise restart



INSTRUCCIONES EXPERIMENTO II

1. Vamos a hacer un experimento adicional, que en términos generales es igual que el anterior, por lo que solo se explicarán las diferencias.

2. Hay dos tipos de participantes: Participantes A y Participantes B. Si durante el experimento anterior eras Participante A o Participante B, este rol se mantendrá a lo largo de todo el experimento.

3. Cada Participante A será aleatoriamente emparejado con un Participante B en cada ronda del experimento. Es decir, la pareja cambiará de forma aleatoria en cada ronda y no volverás a jugar con la misma persona dos veces.

4. El experimento consta de **10 rondas.** Tu pago final será las ganancias en **una de las 10 rondas,** la cual será seleccionada aleatoriamente. Este dinero se sumará a lo que ya ganaste en el experimento anterior.

5. Ten a mano la hoja resumen del experimento anterior, esa información sigue siendo válida.

OK

EXPERIMENT II

1. We are going to do an additional experiment, which in general terms is like the previous one, therefore only the differences will be explained.

2. There are two types of participants: Participants A and Participants B. If during the previous experiment you are Participant A or Participant B, this role will be kept along the whole experiment.

3. Every Participant A will be randomly paired with a Participant B in every round of the experiment. That is to say, the couple will change randomly in every round and you will not play with the same person two times.

4. The experiment consists of 10 rounds. Your final payment will be the earnings in one of the 10 rounds, which will be randomly selected. This money will be added to what you already have at the previous experiment.

5. Have at hand the summary of the previous experiment, because this information is still valid.

Chapter 3

Bargaining and Strategic Moves in the Hawk-dove Game

Abstract: This article studies experimentally Schelling's (1960) strategic moves. In the previous chapter, we captured the effect of threats over the sender's bargaining power, identifying the particular effect of messages, credibility and experience on the final outcome. We follow the same experimental design, now focusing our analysis in the strategy method, following the highlighted advantages of this *cold treatment* for controlling the emotional effect over behavior made by Brandts and Charness (2000). Naturally, we look for *hot-cold* comparisons in all experimental conditions. The new evidence confirms that commitment improves subjects' bargaining power, and credibility is essential for it. Subjects learn gradually the advantages of threatening their counterpart, suggesting that experience is more important in the strategy method. In the same way, credibility is essential for strategic moves, so that the more a sender lies, the less bargaining power he has.

JEL Code: C91, D03, D74

Keywords: Credible threats, negotiation, experiments.

3.1. Introduction

The problem about negotiation on the surplus distribution is not new in the economic literature. Problems such as the division of a fixed amount of money between two persons, and which of them receives the bigger part of the pie have been broadly studied. Schelling's (1956) Essay on Bargaining[1] begins with a basic observation about this problem: it is not a zero-sum game. The players' interests are in strict conflict when we compare any two Pareto-efficient outcomes. But failure of the negotiation is a possible outcome, and when comparing failure to any agreement, the players' interests are in the same direction. Schelling also observes that one bargainer can secure an outcome better for himself implementing a threat. One side can credibly persuade the other to cooperate if he frames that the only other available alternative is the mutually undesired disagreement.

Schelling describes the distributional conflict with an extreme example. When two dynamite trucks meet on a road that is wide enough for only one of them to cross, who backs up? From a tactical point of view, if someone takes the first place, with an aggressive body posture and a dramatic face expression of anger, then bargaining is all over; the one who announce this credible threat is the winner. Authors like Frank (1988), Hirshleifer (1987, 2000) and Elster (1996, 1998) have worked on commitment theory, describing how emotions can be used for achieving credibility[2].

In the previous chapter, we have presented a novel experimental design based on the *hawk-dove game with perfect information*[3]. Given the characteristic of this design, we have recognized that the strategic credibility problem is intrinsically dynamic, however, there are two possible ways for implementing a sequential protocol in the lab. On one hand, Brandts and Charness (2000) define the *hot treatment* when the second player responds to the first player's observed action, that is to play in two-stages. Clearly, the *hot* label has been used to illustrate that sequential

[1] Schelling's (1956) Essay on Bargaining was reprinted as Chapter 2 of The Strategy of Conflict (1960).

[2] The proverb "revenge is a dish best served cold" suggests that vengeance is more satisfying as a considered response enacted when unexpected. The emotional detachment and planning are best for taking revenge. In contrast, the intuition says that threatening is better with "hot blood". We can then rephrase the proverb saying: "threatening is a dish best served hot"

[3] The *hawk-dove game* was studied experimentally by Neugebauer et al. (2008), Duffy and Feltovich (2002) and Bornstein et al. (1997), but their focus were not on bargaining power.

procedures can activate stronger emotional responses. On the other hand, they refer to the "strategy method" as a *cold treatment*, when the second mover decides on a contingent action for each possible first player's move, without first observing it. In order to control for the emotional effect, the focus of this article is the *cold treatment.* Naturally, we keep all other conditions as close as possible, looking for *hot-cold* comparisons[4].

The aim of this article is the strategic communication in a bilateral bargaining. Accordingly with our experimental design, we have new evidence for answering the following questions: *Do people take advantage of commitment opportunity in order to manipulate their opponents' behavior?* and, *is experience in the bargaining process necessary to implement the threat strategy?* Our data set confirm that commitment improves subjects' bargaining power, and credibility is essential for it. Subjects learn gradually the advantages of threatening their counterpart, suggesting that experience is more important in the strategy method.

In addition, we can contribute to the discussion of the behavioral effect of the elicitation procedure, using our join data set. The experimental design allow us to analyze the *hot-cold treatment* effect by two different factors: experience, and communication mechanism[5]. The evidence in Brandts and Charness (2000) shows that the elicitation procedure has no significant impact in subjects behavior. We find the same result in the one-shot game sample, independently of messages and credibility considerations. In contrast, the *hot treatment* has a negative effect in the sender bargaining power in the ten repetition sample. The worst conditions for second players are the real sequential negotiation without messages. The exception is the negotiation with *non-binding messages.* Moreover, subjects lie strategically, but in

[4]In the experimental literature, some hypotheses about the effectiveness of self-serving messages have been tested previously. Binmore et al. (1985) show that outside threats affect bargaining divisions only if divisions themselves make it credible for players to exercise the threat. Güth et al. (1998) and Huck and Muller (2005) studied the positional order effect, and the advantages of burning money in the battle of the sexes game. Güth et al. (2006) studied the strategic behavior of leaders and followers in sequential duopoly experiments with imperfect movement observations. Fischer et al. (2006) and Poulsen and Tan (2007) tested the information transmission in the ultimatum game. More recently, Poulsen and Roos (2010) focused their analysis in a sequential Nash's demand game, and Kimbrough and Sheremeta (2010) studied side-payments for avoiding a lottery contest.

[5]By experience variable, we compare when the game is played in a one-shot game or after 10 period repetitions. By communication mechanism, we compare when subjects negotiate in *without messages*, *non-binding*, or *commitment* treatments.

the *hot treatment* subjects play the actions announced in messages more than in the *cold treatment*. Indeed, the more a sender lies, the less bargaining power he has.

The article is organized as follows. In Section 3.2 the experimental design is described, highlighting the hypotheses on rational behavior. The Section 3.3 presents the main results obtained from the experiment in the *cold treatment*. In Section 3.4 we discus the behavioral effect of the elicitation procedure. Finally, Section 3.5 concludes.

3.2. Experimental Design

The experiment was run with economics and management students at the University of Granada, Spain[6]. The experiment was fully programmed and conducted using the software z-Tree (Fischbacher, 2007). The use of computers allows the instructions to be completely presented on each individual screen, and a brief questionnaire was implemented to ensure their understanding. We conducted a one-shot game, plus a surprising restart with 10 period repetitions. The possibility of playing repeatedly helps us to solve the problem related to the experimental subjects' lack of experience (see Crawford, 1998).

Based on the *hawk-dove game with perfect information*, we recreate a situation where Player A can send a pre-play message, choosing between four different reaction rules, whereas Player B has only two strategies[7]. In order to controlling for the effect of reputation over credibility, we use the absolute strangers protocol. The role of Player A or B is maintained during the whole experiment, but pairs change each period and players never interact with the same partner twice. At the end of each period, participants were completely informed about decisions and payoffs, as repetitions

[6]The *cold treatment* sessions were conducted on May 2009, with a sample of 144 subjects. We conduct additional sessions with the *hot treatment* in November 2010, with 148 subjects more. Nobody participate in two different treatments. Financial support from the Spanish Ministry of Education and Science (grant code SEJ2009-11117/ECON) and the Proyecto de Excelencia (Junta de Andalucía, P07-SEJ-3261).

[7]Treatments allow variations in Player A's set of strategies, and in strict terms the normal form of the *hawk-dove game* was never carried out. However, subjects in the experiment always had the 2x2 payoff matrix, as a reference. The payoffs that appeared in the matrix are in Euros, instead of experimental currency units (ECU). See appendix

looked forth to increase player's experience about the game. Subjects were paid the one-shot game payoffs, plus the payoff of one randomly chosen period out of the 10 periods after the restart. Average payments were 10 Euros.

Our main focus is the possibility of self-committing in a credible way using messages about intentions. Hence, we have three treatments: one without communication; one where subjects can announce their intentions, but messages are cheap talk; and one where subjects can send messages, revealing the action they chose. Based in the strategy method, the characteristics of *without messages* (WM), *non-binding* (NB) and *commitment* (CT) treatments are summarized in the next subsections.

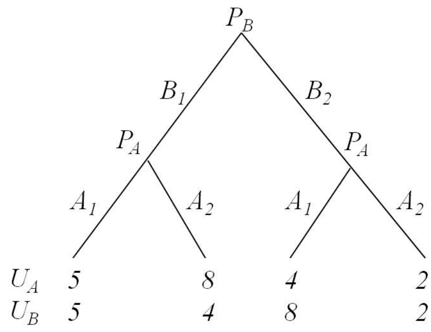
3.2.1. Baseline: Without Messages

There are two types of players: Player A and Player B. In the first stage, Player A chooses a response rule between two actions (A_1 or A_2), for every possible choice of Player B. In other words, Player A decides: *"If Player B chooses* B_1*, I choose* (A_1 *or* A_2) *and if he chooses B2, I choose* (A_1 *or* A_2)*."* In the second stage, Player B must choose between two actions, B_1 or B_2, without any information about Player A's choice.

The strategy method gathers the reaction rules, essential for modeling threats, and removes the decisions sequence. Player A's set of strategies is $X_A = \{A_1A_1, A_1A_2, A_2A_1, A_2A_2\}$, where $x_A = A_{B1}A_{B2}$ represents a possible reaction rule, such that, the first component A_{B1} denotes the action that will be carried out if Player B plays B_1, and the second component A_{B2} the action in case B plays B_2. Player B's set of strategies is $X_B = \{B_1, B_2\}$.

There is imperfect information because no player knows his opponent's decision at the choosing moment. Given that there are simultaneous moves, this game is the reduced normal form of the *hawk-dove game with perfect information* in Figure 3.1. The theoretical prediction by backward induction is that Player A chooses, according to his incentives, the strategy A_2A_1. Strictly speeching, Player A chooses A_2 conditioned to B_1, and A_1 conditioned to B_2, because this is his best response in each subgame. Player B plays B_2 because it is his best response if he forecasts the opponent behavior, reaching the outcome $(4, 8)$.

Figure 3.1: Hawk-dove game with perfect information



3.2.2. Non-binding Messages

In this treatment, decisions are the same as in the previous one. The difference is that Player A has the possibility to send a payoff irrelevant message about his intention of play. If Player A wants to send a message, he must choose one, and it is clear for both players that it will not necessarily coincide with the actions that define the final outcome. If Player A *"does not want to send a message"*, Player B will receive the announcement *"there is no message from Player A,"*. In other case, when A chooses to send a message, his decision consists on choosing between two actions (A_1 or A_2), for every possible choice of Player B. In other words, Player A chooses both a response rule and a message like this: *"If Player B chooses $B1$, I choose (A_1 or A_2) and if he chooses $B2$, I choose ($A1$ or $A2$)"*.

In summary, in Stage 1 Player A chooses a response rule, and next he chooses a message, with the same response rule structure. In Stage 2, Player B observes the message, and he chooses a single action. It is common knowledge that response rules and messages do not necessarily coincide and final payoffs are calculated using actions, not messages. The fact that messages do not modify the game structure implies that Player A is going to be indifferent between any of them. They are always playing the same extensive game, and by backward induction, there is not evident advantage in using this type of communication.

The theoretical prediction is the same as in the WM treatment: Player A chooses strategy A_2A_1, and Player B chooses strategy B_2 independently of receiving a message or not, reaching the outcome $(4, 8)$. We label the message A_2A_1 as *warning*, using Schelling's classification, because that is Player A's strategy in the Subgame

Perfect Nash Equilibria.

3.2.3. Commitment Treatment

In this treatment, the difference is that Player A has the possibility of sending a message to Player B, revealing the action he chose. Thus, if Player A sends a message, in the second stage Player B faces the choice of his opponent before making the decision. There is no possibility for lying. If Player A decides that *"he does not want"* to send a message, Player B will receive the announcement *"there is no message from Player A,"* and his incentives are the same as in WM treatment.

Player A chooses whether he wants to send a commitment message or not. This decision is essential for testing if communication is useful to increase the sender's bargaining power. In this sense, Player A now has the possibility to play first, committing himself with a reaction rule, and Player B knows it. The theoretical prediction is that Player A is going to play the strategy A_2A_2, and he is going to reveal his decision. We label the message A_2A_2 as a *threat* based on an aggressive statement that intends to deter the opponent to behave aggressively, too. Faced with this binding threat, Player B's best response is choosing strategy B_1 in order to stay away from the outcome $(2, 2)$. In this treatment, we find an essential result in relation to commitment theory, where Player A changes the unfavorable situation to reach the outcome $(8, 4)$, which is his larger available payoff.

There is another interesting strategy, the *promise* A_1A_2. Player A might stipulate that he will play reciprocally, cooperating if other one does it, but clarifying that the aggressive behavior will be punished in an equal way. This message has the characteristics of a threat, but we prefer the label *promise* because it establishes to reward the cooperation with a not best response action. Though this strategy has serious problems of credibility, it has the advantage of eliminating the conflict on the surplus distribution. Player B must choose between the equitable results $(5, 5)$ or $(2, 2)$.

The characteristics of WM, NB, and CT treatments are summarized in Table 3.1. Note that Player B's set of strategies and the set of final outcomes are constant among treatments. The changes between treatments are Player A's set of strategies, including the sending message possibility, and Player B's information sets.

Table 3.1: Timing in the different treatments

	WM	NB	CT
Stage 1	Player A Response rule	Player A Response rule Message	Player A Response rule Reveling actions
Stage 2	Player B	Player B	Player B
Theoretical Prediction	(A_2A_1, B_2) $(4, 8)$	(A_2A_1, B_2) $(4, 8)$	(A_2A_2, B_1) $(8, 4)$
Sample	30 subjects	56 subjects	58 subjects

3.2.4. Theoretical Considerations and Hypothesis

Based on the commitment theory, Player A's bargaining power increases if he is able to threat with an aggressive arrangement, and makes it clear that this is the unique relevant option. This strategic move can be implemented by the possibility of sending binding messages. In consequence, the experimental design recreates a situation where sending fully credible messages is available for Player A. With this asymmetry in communication, we can expect that Player A uses the option of sending the *threat* message because it is the way of turning the outcome into his favor.

Following this argument, Player A's largest payoffs would be reached in CT. The theoretical prediction is to reach the $(8, 4)$ outcome with *commitment messages*, and the $(4, 8)$ otherwise, because *without messages* or *non-binding messages* have the same properties in strategic terms. In the same way, the average total payoff ($U_A + U_B$) will show the importance of the messages to solve the coordination problem, because for both players it is rational to avoid the $(2, 2)$ outcome. In the game that we are using in this experiment, total payoffs in equilibrium are 12 € independently of the treatment. We capture this predictions in Hypotheses 3.1, 3.2 and 3.3, as it appears in Table 3.2.

Table 3.2: Theoretical predictions

Do people take advantage of commitment opportunity in order to manipulate their opponents' behavior?
Hypothesis 3.1 *Player A's payoffs are equal in WM and NB treatments.*

Hypothesis 3.2 *Player A's payoffs are larger in CT than in NB treatment.*

Is the outcome efficient under the different communication mechanisms?
Hypothesis 3.3 *The average total payoff is not different between treatments.*

Would people lie strategically to turn the agreement into their favor?
Hypothesis 3.4 *In NB treatment, Player A chooses the response rule warning, but he sends the threat message.*

Is experience in the bargaining process necessary to implement the threat strategy?
Hypothesis 3.5 *The final outcome achieved does not change when game repetitions are going on.*

Does elicitation procedure have a behavioral effect?
Hypothesis 3.6 *There are not differences in behavior between hot and cold treatments.*

Hypothesis 3.4 is related with messages' credibility. In NB treatment Player A can send messages, which can be used for bluffing. The theoretical prediction is Player A chooses the action *warning*, nevertheless, he can choose a different message in order to induce the opponent to choose B_1. This is exactly the purpose of threatening.

The theoretical predictions are based on backward induction reasoning in the *hawk-dove game with perfect information*. Although we consider this hypothesis as a good reference point, from a behavioral point of view (Croson et al., 2003), the learning process is very important to show how decisions change as experience increases. Hypothesis 3.5 is based in the usual assumptions of the rational choice theory. We must expect that behavior is not time-dependent. Finally, we do not expect any difference between *hot* and *cold treatments*, as it was found in Brandts and Charness (2000).

3.3. Results in the Cold Treatment

This experiment seeks to stand out if messages, credibility, and experience are essential to improve Player A's bargaining power. To achieve this aim, in Section 3.3.1 we compare the outcome distribution in the one-shot game (OSG) and in the last repetition. We analyze the outcomes progression over periods in Section 3.3.2. Finally, in Section 3.3.3 we run a logit model to understand the variables that influence Player B to choose B_1.

3.3.1. One-shot Game vs. Last Repetition

The average Player A's payoffs in the OSG were 5.5 € in WM, 5.5 € in NB, and 5.9 € in CT (see Table 3.3). It is clear that payments reached in CT are not significantly different with respect to other treatments (WM and NB, p = 0.955; NB and CT, p = 0.406)[8]. With regard to efficiency, the average total payoff is 10.4 € in WM, 10.6 € in NB, and 11.6 € in CT. The highest payoff obtained in the CT treatment is not significantly different with respect to the other treatments (WM and NB, p = 0.864; NB and CT, p = 0.347).

Table 3.3: Payoffs distribution

	One-shot Game			Last Repetition (Period 10)		
Outcome	WM	NB	CT	WM	NB	CT
(2,2)	20	18	3	60	14	7
(4,8)	33	36	42	20	68	7
(5,5)	0	0	7	0	4	24
(8,4)	47	46	48	20	14	62
Average	(5.5,4.9)	(5.5,5.1)	(5.9,5.7)	(3.6,3.6)	(4.3,6.5)	(6.6,4.4)

Payoffs distribution are in percentage. Average values are in Euros.

As shown in Table 3.3, the evidence in the last repetition supports Hypothesis 3.2 about A players' bargaining power. Average Player A's payoffs in Period 10 were 3.6 € in WM, 4.3 € in NB, and 6.6 € in CT. Differences of CT respecting others treatments are significant at 1% (WM and NB, p = 0.040; NB and CT, $p < 0.01$). With regard to efficiency, the average total payoff is 7.2 € in WM, 10.8 € in NB, and 11.0 € in CT. The difference between WM and NB is significant at 1%, but the

[8]Mann-Whitney test, two-sided. Between subjects comparisons.

difference between NB and CT is not significant (WM and NB, $p < 0.01$; NB and CT, $p < 0.396$).

In the OSG, differences between treatments are not significant. Our conclusion is that played strategies and outcomes depend neither on the possibility of sending messages nor on credibility, if the interaction is carried out only once. However, in the last repetition the findings are clearly different. The theoretical prediction is very salient in CT and NB, with the outcomes $(8, 4)$ and $(4, 8)$ being reached in the 62% and 68%, respectively. However, the great disagreement rate in WM is surprising, 60% of subjects end up in the $(2, 2)$ outcome. In the last period, evidence supports Hypothesis 3.2, but we must reject Hypotheses 3.1 and 3.3. This findings are summarized in Result 3.1

Result 3.1 *Player A does not have more bargaining power in CT if the game is played only once. In contrast, the behavior after 10 repetitions is in favor of this hypothesis. Furthermore, there is a great disagreement rate in WM.*

In the last repetition, the evidence shows the logical treatments order, in terms of Player A's payoff. Player A turns payoffs into his favor in CT. The question is how does Player A reaches the highest share of the pie? The answer goes in the direction of the theoretical predictions, threatening in a credible way about falling down into the worst result if his rival does not cooperate.

Although we are going to study in more detail the subjects behavior in the next subsection, the descriptive statistics in Table 3.4 is a good general picture for treatments comparisons. First, in WT the variable B_1 is the proportion of B Players choosing to cooperate. Second, in NB we present the distribution of actions and messages. In this case, the conditional proportion B_1 is a measure of the message effectiveness, and *Lie* is the proportion of A players that choose an action different than the ones in the message. Finally, in CT the interpretation of Message and B_1 is similar.

Player A in NB is sending the message because his situation would be even worse if he chose not to do so. The conditional proportion of Players B choosing B_1, given no-message, is only 21%. In general, A players has more bargaining power if they choose *threat* or *promise*. However, individuals' behavior in NB has particular characteristics. The data allow us to answer the following questions: Are people strategic

Table 3.4: Message Effectiveness

	WM		NB				CT	
	B_1	Action	Message	B_1	*Lie*		Message	B_1
A_1A_1		3	4	18	100		0	0
Promise		11	41	26	83		16	98
Warning		66	6	12	47		1	0
Threat		20	25	32	72		56	76
No-Mess	32		24	21			27	25
Total	32	100	100	25	78		100	65

Average from Period 1 to 10. Values are in percentage.

liars? Are subjects naive believers of these messages? Is credibility important for commitment?

Senders are choosing those messages for inducing Player B to play B_1 (25% *threat* and 41% *promise*). The problem comes when we consider that some of them are strategic liars. Those subjects who send the *threat* and *promise* message are lying at 72% of the times and 83%, respectively. The 66% of the subjects are playing the *warning* strategy. There are strategic liars, but subjects are not naive believers. The strategic move look for influence the other player's decision, but without credibility, they reach the opposite outcome. We cannot reject Hypothesis 3.4.

Result 3.2 *There is huge evidence of strategic lying behavior. Senders are choosing those messages for inducing Player B to play B_1, but in the majority of cases they choose the warning strategy.*

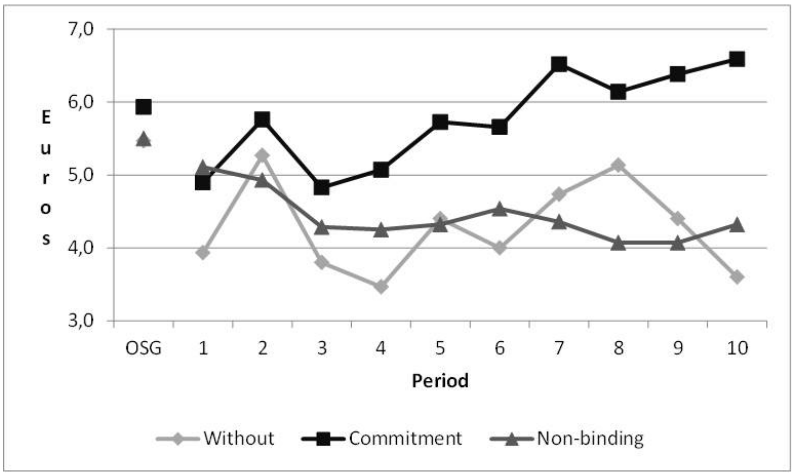
In CT, 56% and 16% of subjects are playing the *threat* and *promise* messages, respectively. Clearly, the message in CT is the way to inform about the *threat*, and Player B must choose between the outcomes $(8, 4)$ or $(2, 2)$. Facing this commitment message, 76% of subjects behave like a dove. For *promise* the analysis is similar. Although the effectiveness of *promise* is the highest (98%), this strategy is less profitable given that Player B must choose between the outcomes $(5, 5)$ or $(2, 2)$.

As previously stated, Result 3.1 suggests that outcomes and strategies seem to be not time-independent. For this reason, the following sections looks for disentangling the learning process, if any.

3.3.2. Outcome Progression

When the 10 periods are analyzed, Player A's average payoffs are of 4.3 € in WM and 4.4 € in NB, and it is possible to state that the trends are flat or negative. However, in CT, Player A's average payoffs have a positive trend. For the first 5 periods, the average payments are 5.3 €, and for Period 6 to 10 they rise to 6.3 €, as Figure 3.2 shows. For Periods 1 to 10 this difference between NB and CT treatment is significant at 1% (WM and NB, $p = 0.521$; NB and CT, $p < 0.001$)[9].

Figure 3.2: Average Player A's payoffs



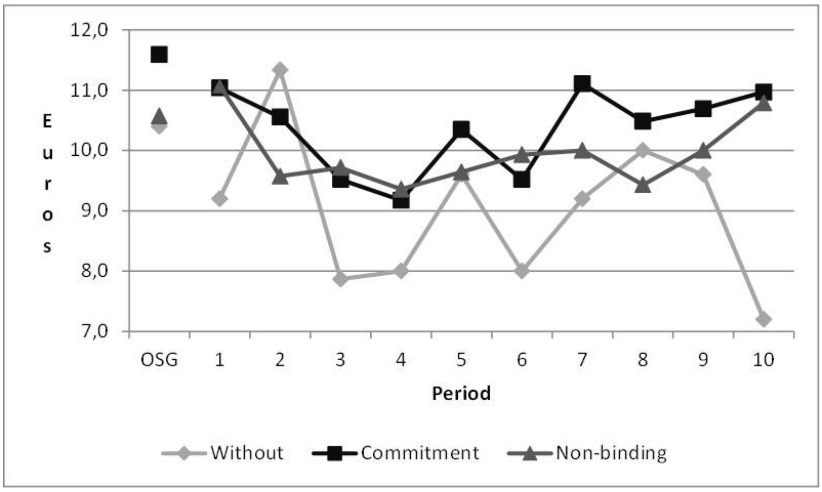
Result 3.3 *In the aggregate, Player A's payoffs are statistically higher in CT. As experience increases, A players gradually reach a higher bargaining power.*

In Figure 3.3 we show that the bargaining process imply an efficiency loss, given that the theoretical prediction is 12 € in all treatments. However, it is possible to conclude that messages help coordination, because total payoffs reached in CT and NB are higher with regard to WM. For Periods 1 to 10, the average total payoff is 9.0 € in WM, 10.0 € in NB and 10.3 € in CT. The difference is significant between WM and NB (WM and NB, $p < 0.001$; NB and CT, $p = 0.585$). Notice that the worst situation is WM in the last period, as previously stated.

In summary, we disentangle experimentally the effect of messages and its credibility to improve Player A's bargaining power. Nevertheless, messages can also be used

[9]Mann-Whitney test, two-sided. In this section we take the average for the 10 periods repetitions. Between subjects comparisons.

Figure 3.3: Average total payoffs ($U_A + U_B$)



to avoid the outcome (2, 2), which is a consequence of the coordination problem inherent to the *hawk-dove game*. It is clear that Player A increases his bargaining power in CT, however, it is surprising that with experience the outcome in NB is not only in favor of Player B, but it is also his best scenario. There are many strategic liars, destroying their own credibility. In addition, the worst condition for both players is the negotiation without communication, because the path goes far from the efficiency frontier.

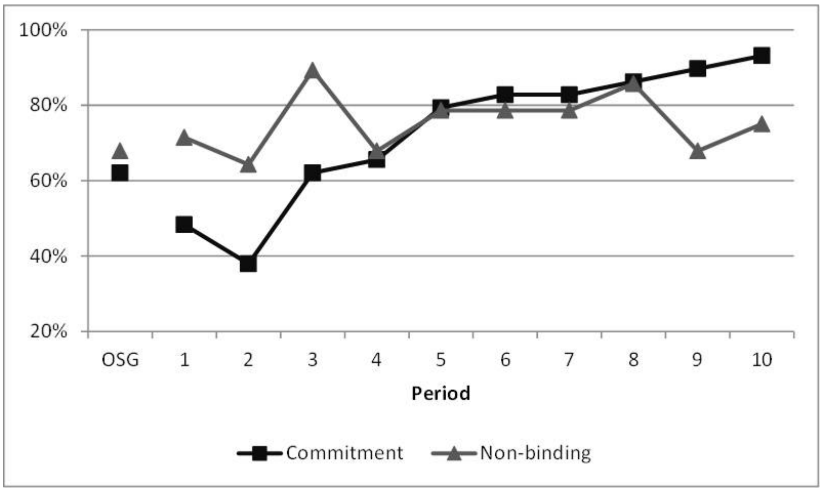
The strategy method implies an interaction with imperfect information. In WM, subjects must play without any information about the opponent decision. As in Roth (1985), the disagreement rate is high only as a consequence of coordination among multiple focal points. If a player is changing his choice, any of the equilibrium outcomes is selected, and then a disagreement point is reached in a proportion higher than in the other treatments. We conclude that the Subgame Perfect Nash Equilibrium is not an accurate prediction in WM.

Result 3.4 *The strategy method implies simultaneous moves in WM. This characteristic leads to a negative effect on outcome efficiency. Total payoffs are statistically equal in NB and CT, but in WM this variable is significantly low.*

In the next section we are going to analyze in detail the strategic role of communication. For this reason, we must highlight another interesting finding. There are no significant differences in the proportion of sent messages between NB and CT (p

= 0.42)[10] when we analyze the 10 periods sample. It indicates that messages are considered to be useful independently of their credibility. However, for the first 5 periods, there is a significant difference in the proportion of sent messages, given that in NB the proportion is 74% and in CT it is 58% ($p < 0.01$). For Periods 6 to 10, the messages were used in 77% of the times in NB and in 86% in CT. This difference is significant, but in an opposite direction than in the previous result (p = 0.031). This evidence suggests that the decision of sending a message is more difficult at the beginning of CT, but with experience this proportion grows up, as it appears in Figure 3.4.

Figure 3.4: Proportion of sent message



3.3.3. Strategic role of communication

For a deeper analysis, we study Player B's decision as an indicator of Player A's bargaining power. Table 3.5 presents the results of 5 panel logit regressions of Player B's decision to play B_1, including dummies variables for *message*, CT and NB, a linear trend (*Period*), the time that Player B takes to make his decision (*choosing time*) as regressors. Regressions 2 and 3 include a dummy for the OSG, whereas Regressions 4 and 5 include dummies of Player B's payoff in $t-1$, and the dummy variable Lie_{t-1}, for capturing the fact that Player B could face a lie in the previous

[10] Given that sent messages is a binary variable, we use the non-parametrical test of proportions, two-sided.

period. Regressions 3 and 5 also include the interaction between *period* and each one of the treatments dummies[11].

Table 3.5: Regression analysis explaining Player B's decision

Player B (B_1)	(1)	(2)	(3)	(4)	(5)
Message	1.483***	1.414***	1.185***	1.877***	1.650***
	(0.247)	(0.259)	(0.266)	(0.300)	(0.309)
OSG		0.662*	0.627*		
		(0.350)	(0.341)		
CT	0.380	0.413	-0.615	0.214	-0.697
	(0.453)	(0.456)	(0.639)	(0.499)	(0.791)
NB	-1.718***	-1.692***	-0.595	-1.971***	-0.760
	(0.480)	(0.482)	(0.664)	(0.565)	(0.833)
Period		0.056	0.039	0.041	0.048
		(0.036)	(0.061)	(0.039)	(0.071)
Period·CT			0.204***		0.179*
			(0.076)		(0.093)
Period·NB			-0.164**		-0.193**
			(0.077)		(0.094)
$\text{Payoff}_{t-1} = 8$				-0.632**	-0.582**
				(0.282)	(0.291)
$\text{Payoff}_{t-1} = 5$				0.085	-0.147
				(0.389)	(0.402)
$\text{Payoff}_{t-1} = 4$				-0.057	-0.274
				(0.286)	(0.295)
Lie_{t-1}				-0.251	-0.056
				(0.345)	(0.357)
Choosing time	0.007	0.009	0.011*	0.008	0.011
	(0.005)	(0.006)	(0.006)	(0.007)	(0.008)
N	792	792	792	720	720

Logit panel data estimation with random effects. Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notice that regressions in Table 3.5 include the variables that influence the probability of choosing B_1, turning the outcome in favor of Player A. The possibility of sending messages is a persuasive strong mechanism. The *message* coefficient is always positive, and we can conclude that this effect is significantly robust for explaining the high rates of Players A choosing to send messages. The OSG coefficient is positive and significant. We can conclude that at the beginning the theoretical predictions are not accurate. In addition, the interaction between *period* and CT suggest that subjects learn gradually how to use the message properly.

[11] We also estimated logit regressions clustering by individual to compute the correct statistical significance for these interaction terms (Norton et al., 2004). Notwithstanding, the significance of these interaction terms remains under this specification.

In contrast, Table 3.5 shows a significant negative effect of NB, and the learning process goes in favor of Player B. The interaction between *period* and NB is negative and significant. This result goes in line with the evidence from previous literature (Croson et al., 2003; Brandts and Charness, 2003), where the willingness to punish an unfair action is triggered by a deceptive message. In our case, a deceptive message inclines Player B's decision towards strategy B_2. Lies destroy the positive reciprocity and lead outcomes to be closer to those that the rational choice theory predicts. In the *hawk-dove game*, to punish is simply to play in an aggressive way, which leads to reaching the Subgame Perfect Nash Equilibrium.

Result 3.5 *The behavior of those subjects who face the message explains variation between treatments. The evidence suggests that the decision of sending a message is more difficult in CT at the beginning, but with experience this proportion grows up.*

In order to go one step ahead in this analysis, it is necessary to understand Player B's decision, conditioned to facing a message. In Table 3.6, we perform a regression analysis only for those B Players who faced the message, including dummies of the message sent by Player A as regressors. The empirical evidence shows that Player A needs a learning process to understand the strategic role of communication. As we could see above, at the beginning, subjects were not using very often the message in CT, but this situation changes when they learn how to use it properly. This fact is captured in the interaction between *period* and CT, with a positive effect.

The *threat* and *promise* coefficient are positive and significant. With experience, Player A learns the importance of the *threat* in CT. For Periods 1 to 5 this strategy is played by 49.7% of subjects who send the message, and it rises to 71% in the last 5 periods. But this message is not the only one to increase Player A's bargaining power. Although in a minor proportion (16% for Periods 1 to 10), the strategy *promise* is also used to induce Player's B decision[12].

Tables 3.5 and 3.6 contain variables that suggest an explanation about the way in which the learning process works: $\text{Payoff}_{t-1} = 8$, Lie_{t-1} and choosing time. When Player B reaches the 8 € payoff in a previous period, this fact reduces the probability to play B_1 in the current period, compared with reaching the 2 € payoff. This

[12]This result is particularly interesting because the strategy *promise* is dominated by *warning*, but can be considered a safe way for obtaining a satisfactory outcome of $(5, 5)$, avoiding the conflict about outcome distribution.

Table 3.6: Regression analysis explaining Player B's decision, conditional to *message* = 1

Player B (B_1)	(1)	(2)	(3)	(4)	(5)
Promise	2.776***	2.942***	2.210**	2.036**	1.710*
	(0.982)	(1.004)	(0.940)	(0.993)	(0.985)
Warning	-0.837	-0.839	-0.798	-1.200	-1.443
	(1.154)	(1.178)	(1.100)	(1.316)	(1.298)
Threat	2.456**	2.617***	1.736*	1.583	1.135
	(0.972)	(0.999)	(0.938)	(0.993)	(0.987)
OSG		0.737	0.404		
		(0.583)	(0.553)		
CT	2.840***	2.829***	0.272	2.927***	0.563
	(0.444)	(0.448)	(0.687)	(0.593)	(0.874)
Period		0.070	-0.115*	0.083	-0.094
		(0.054)	(0.069)	(0.060)	(0.078)
Period·CT			0.436***		0.425***
			(0.095)		(0.118)
$\text{Payoff}_{t-1} = 8$				-0.871**	-0.838*
				(0.420)	(0.444)
$\text{Payoff}_{t-1} = 5$				-0.137	-0.566
				(0.608)	(0.645)
$\text{Payoff}_{t-1} = 4$				0.076	-0.269
				(0.426)	(0.452)
Lie_{t-1}				-0.291	-0.120
				(0.414)	(0.428)
Choosing time	0.013*	0.016*	0.018**	0.020*	0.025**
	(0.007)	(0.008)	(0.008)	(0.010)	(0.011)
N		460		423	

Logit panel data estimation with random effects. Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

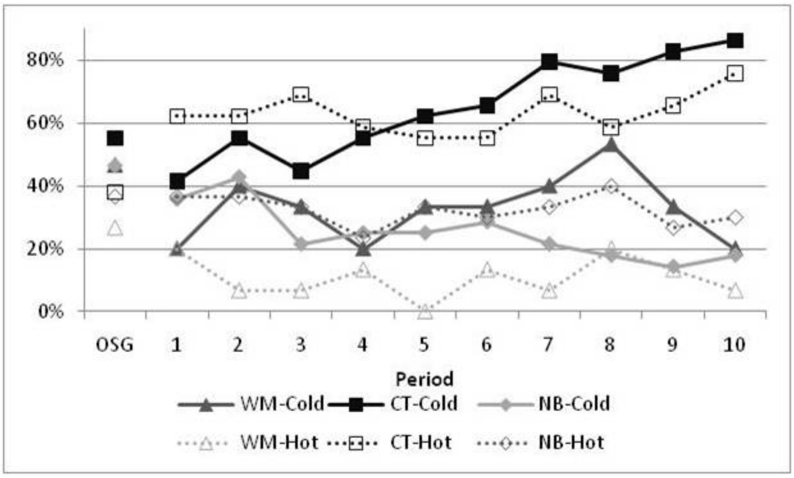
coefficient is always negative and significant. In the same way, Lie_{t-1} coefficient is always negative, but not significant. If Player B faces a lie in a previous period, then the probability of cooperating in the current period is reduced. The variable *choosing time* refers to the seconds spent by Player B in order to make his choice. This coefficient is always positive and conditional to message = 1 it is significant. This fact suggests that choosing B_1 is cognitively more demanding for Player B, because it is necessary to spend more time evaluating the consequences.

3.4. Hot vs. Cold

Brandts and Charness (2000) analyze whether responses depend on the elicitation procedure. They argue that, due to its simplicity, *prisoner's dilemma* and *hawk-dove game* are a good starting point to disentangle the *hot-cold* problem form issues of complexity. Their results show no difference in behavior between the two treatments, and also find evidence of the stability of subjects' preferences with respect to their behavior over time. Clearly, our design it is not a direct replication of their experimental protocol, however we can contribute to this discussion comparing these treatment effects, using the join data set in our experiments.

For a complete analysis of behavior, the best variable is the proportion of subjects playing B_1. In order to simplify the notation we will use the abbreviation of every treatment, which is, WM, NB and CT, differentiating between *hot* and *cold*.

Figure 3.5: Proportion of Player B playing B_1, by treatments



As appear in Figure 3.5, in the OSG 47.7% of B players choose B_1 in WM-Cold and 26.7% in WM-Hot. Although cooperation is higher in the *cold treatment*, this difference is not significant ($p = 0.255$)[13]. We can note the same observation in NB and CT, given that this proportion is 46.4% in NB-Cold and 36.7% in NB-Hot ($p = 0.457$); and 55.2% in CT-Cold and 37.9% in CT-Hot ($p = 0.188$). Clearly, our data support the Hypothesis 3.6, preferences do not depend on the elicitation procedure in a single-period decision protocol. However, we must recognize that the strategic method has a qualitative effect, for this reason, our game repetitions could be a good way for establish whether this effect is systematic or not.

[13]Given that B_1 is a binary variable, we use the non-parametrical test of proportions, two-sided.

Result 3.6 *If we consider the OSG sample, Player B cooperates more in cold than in hot treatment. However, this difference is not statistically significant.*

There is a salient observation in the 10 period sample, B players behavior is the most aggressive in the WM-hot treatment (see Figure 3.5). Player B's cooperation is 10.7% WM-Hot, the lower average cooperation, and the difference with respect in WM-Cold is significant (32.7% in WM-Cold, $p < 0.001$). Therefore, the sequential game without messages is best environment for B players, in fact, their average payoff is 6.54 € for the 10 periods repetition. In CT, the strategy method has a negative effect over Player B' cooperation. However, we do not find great *hot-cold* differences. B Players' cooperation is 64.8% CT-Cold and 63.1% in CT-Hot ($p = 0.665$). On the contrary, we must highlight that cheap talk threatening seems to be a dish best served hot, since B_1 is played at 25.0% in NB-Cold and 32.3% in NB-Hot ($p = 0.051$).

Result 3.7 *When subjects' experience increases, the best scenario for Player B's bargaining power is WM-Hot, and the worst is CT-Cold.*

With the commitment theory we can understand these results. Clearly, the simplest strategic move is to play first, making public that B_2 strategy has been chosen. In this situation, the bargaining power is on Player B's side. Why this tactical advantage does not appear in the strategy method? Because this procedure implies a making decision procedure without any information about others decisions. On the other hand, a threat is a second mover message, looking for change their unfavorable position. We have been shown that this reaction strategy is successful when Player A can communicate their play. In fact, in CT Player A has the first mover advantage.

The result in the *non-binding treatment* is more difficult to understand. The B players' cooperation rate is in between the others treatments, but the *hot-cold* effect is not obvious. For a deeper analysis, Table 3.7 presents the results of 6 panel logit regressions, now by treatments. The dummy variable *hot treatment* reinforce the previous observation in WM and CT, given that is negative and significant. For NB, we run two different regressions, in order to incorporate the Lie_{t-1} variable. In this case, we can see that elicitation procedure is not the important variable for explaining Player B's behavior, the possibility to send messages and their credibility are the important ones. A message is credible if Player B believes that it is highly probable to face a truth-telling message. In this way, A Players loose part of their bargaining power when commitment is not so strong. Player B faces a Lie 56.4% of the times in NB-Cold and 33.3% in NB-Hot, and this difference is significant at 1% ($p < 0.01$).

Table 3.7: Regression analysis explaining Player B's decision, by treatments

Player B (B1)	WM	CT	NB		The whole sample	
Hot treatment	-1.644***	-1.267***	0.273	0.317	-0.668***	-0.764***
	(0.606)	(0.331)	(0.430)	(0.498)	(0.245)	(0.278)
Message		2.053***	0.637**	0.648**	1.314***	1.677***
		(0.302)	(0.284)	(0.311)	(0.207)	(0.235)
Final decision		1.100***			1.041***	1.282***
		(0.285)			(0.281)	(0.313)
Commitment treatment					-0.192	0.178
					(0.484)	(0.598)
Non-binding treatment					-0.386	-0.100
					(0.495)	(0.614)
One-shot game	0.718	-0.395	0.120		0.121	
	(0.611)	(0.371)	(0.393)		(0.243)	
Period	0.074	0.045	-0.065*	-0.051	-0.010	0.035
	(0.060)	(0.039)	(0.039)	(0.042)	(0.048)	(0.056)
Period·Commitment					0.134**	0.057
					(0.056)	(0.068)
Period·Non-binding					-0.068	-0.108
					(0.057)	(0.069)
Lie_{t-1}				-0.465*		-0.552**
				(0.249)		(0.244)
Choosing time	0.043**	-0.002	0.017***	0.022***	0.010***	0.010**
	(0.018)	(0.006)	(0.006)	(0.007)	(0.004)	(0.005)
N	330	638	638	580	1606	1460

Logit panel data estimation with random effects. Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

There are two complementary explanations for the higher rate of lying in NB-Cold: the emotional effect and the observability of actions and messages. The *hot* environment triggers stronger emotional responses, because subjects could feel disappointed for the aggressive behavior of their counterpart. As previously stated, authors like Frank (1988), Hirshleifer (1987, 2000) and Elster (1996, 1998) have shown that emotions like anger or gratefulness can be used for achieving credibility, in threats and promises respectively.

There is a complementary explanation, as a consequence of the experimental design. Participants were completely informed about decisions and payoffs, but this information is not the same in each treatment. The strategy method reveal more lies because B players face the complete response rule, not only the real payoff relevant decision. This fact is very salient in threats. When the threat is successful inducing Player B to cooperate, then for Player A it is completely rational to tell the truth. On the contrary, credibility is tested when Player B is an aggressive negotiator, leading to the worst outcome if Player A choose according with the message.

Result 3.8 *Subjects lie strategically, but this behavior is more salient in the cold treatment.*

3.5. Conclusions

In bargaining theory, *strategic moves* are actions taken prior to playing a subsequent game, with the aim of modifying the available strategies, information structure or payoff functions. The idea is to change the opponent's beliefs, by making it credible that the position is unchangeable. For this aim, we implement a novel experimental design based on the *hawk-dove game with perfect information.* The advantage of this design is that it allows us to have clear theoretical predictions, highlighting the particular effect of messages, credibility and experience on the sender's bargaining power.

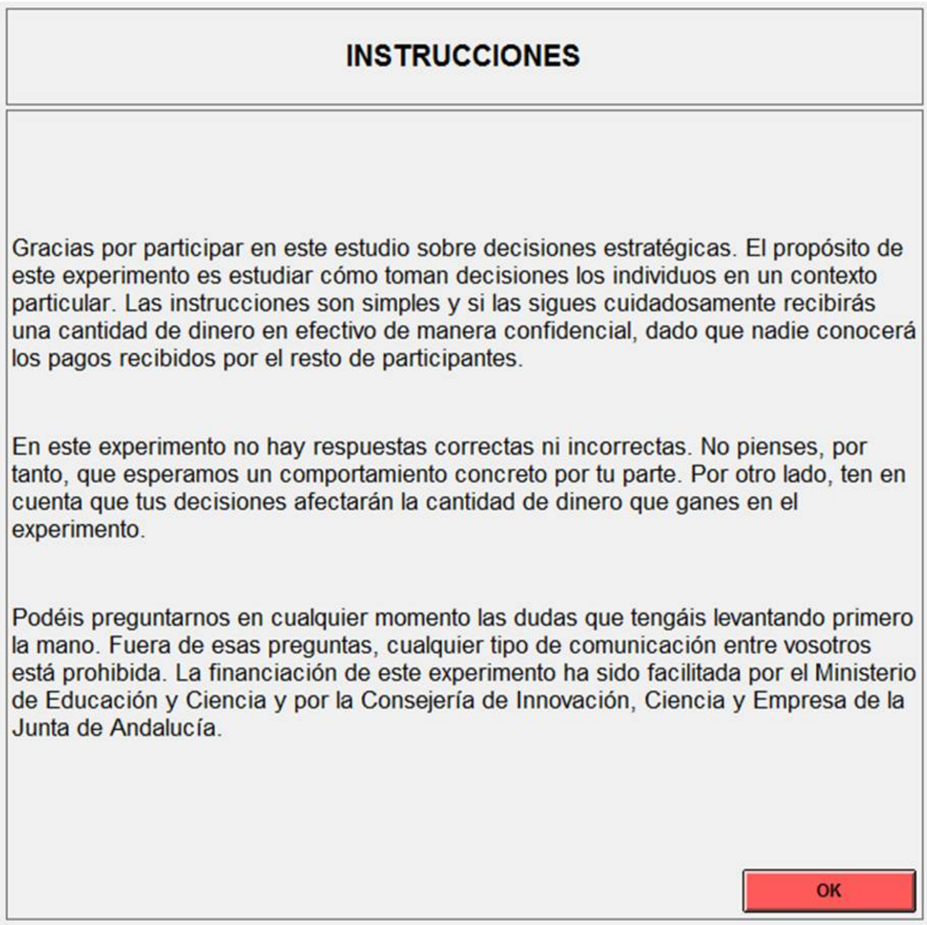
In the *hawk-dove game*, a threat is an aggressive statement, increasing the risk for both players, in order to deter the opponent from behaving aggressively, as well. The experimental analysis shows that commitment improves subjects' bargaining power, as it is predicted theoretically. Nevertheless, experience is more important than expected, because it is necessary to threaten the counterpart, and it seems that this decision is not easy to make. Credibility is also essential. There is evidence of strategic lying behavior, but when subjects tell the truth in a higher proportion, they increase their bargaining power.

We base our predictions in the Subgame Perfect Equilibrium concept. Nevertheless, this conceptualization does not give us clear predictions in case of cheap talk. In this way, our experimental subjects do not choose the non-binding messages randomly, they send those messages that induce the other part to cooperate. Clearly, when subjects lie more, the message credibility is lower. This result leaves us to study in more detail the theoretical properties of non-binding messages in bargaining environments.

In future research, our modeling will deepen in the credibility condition. Binding messages imply a degree of commitment at a 100%, but this condition is very restrictive. In a practical situation, people can believe in a message if it is highly probable that the sender is telling the truth, defining the credibility as a continuum variable, not a binary one. This characteristic could open the possibility of studying commitment with endogenous mechanisms of credibility, taking into consideration the different ways to obtain reputation.

Appendix: Instructions in the Cold Treatment

Figure 3.6: Screen 1: Common instructions

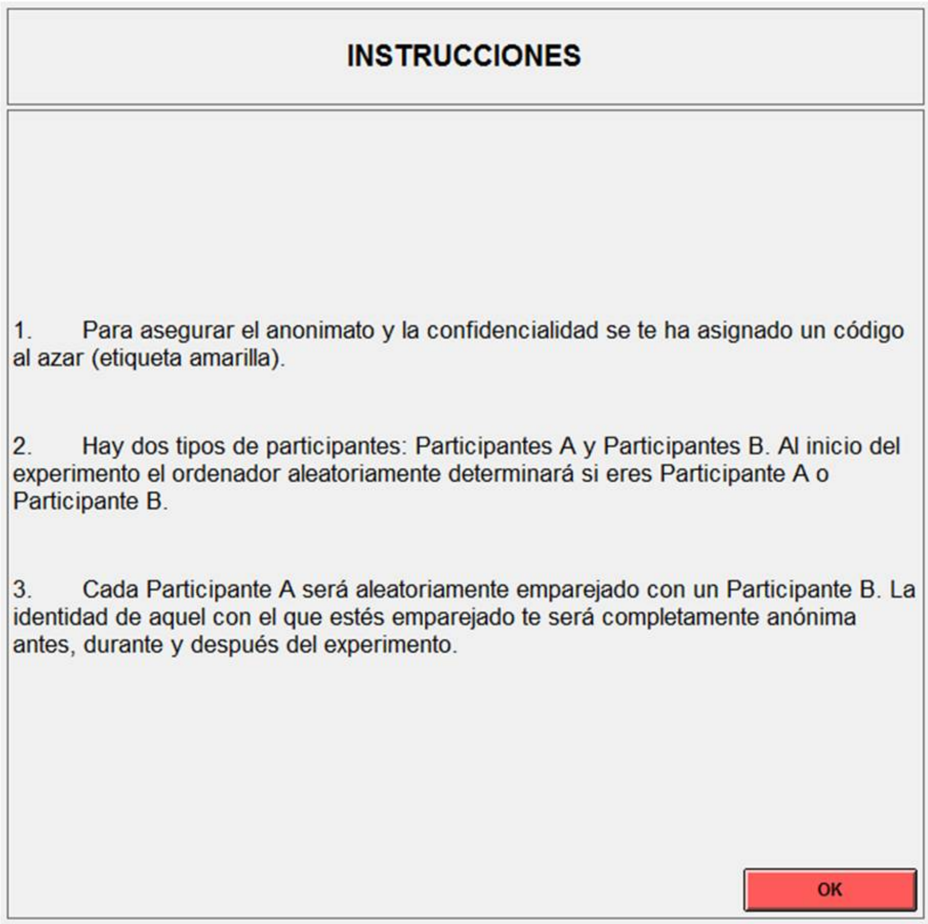


Thank you for taking part in this study on strategic decisions. The intention of this experiment is to study how the individuals make decisions in a particular context. The instructions are simple and if you follow them carefully you will receive an amount of money in a confidential way, so that nobody will know the payments that the other participants will receive.

In this experiment there are neather correct nor incorrect answers. Do not think, therefore, that we expect a concrete behavior for your part. On the other hand, bear in mind that your decisions will affect the quantity of money that you earn in the experiment.

You can ask us at any time about the doubts that you have by raising your hand. Except of these questions, any type of communication among you is prohibited. The financing of this experiment has been facilitated by Ministerio de Educación y Ciencia and by the Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía.

Figure 3.7: Screen 2: Common instructions

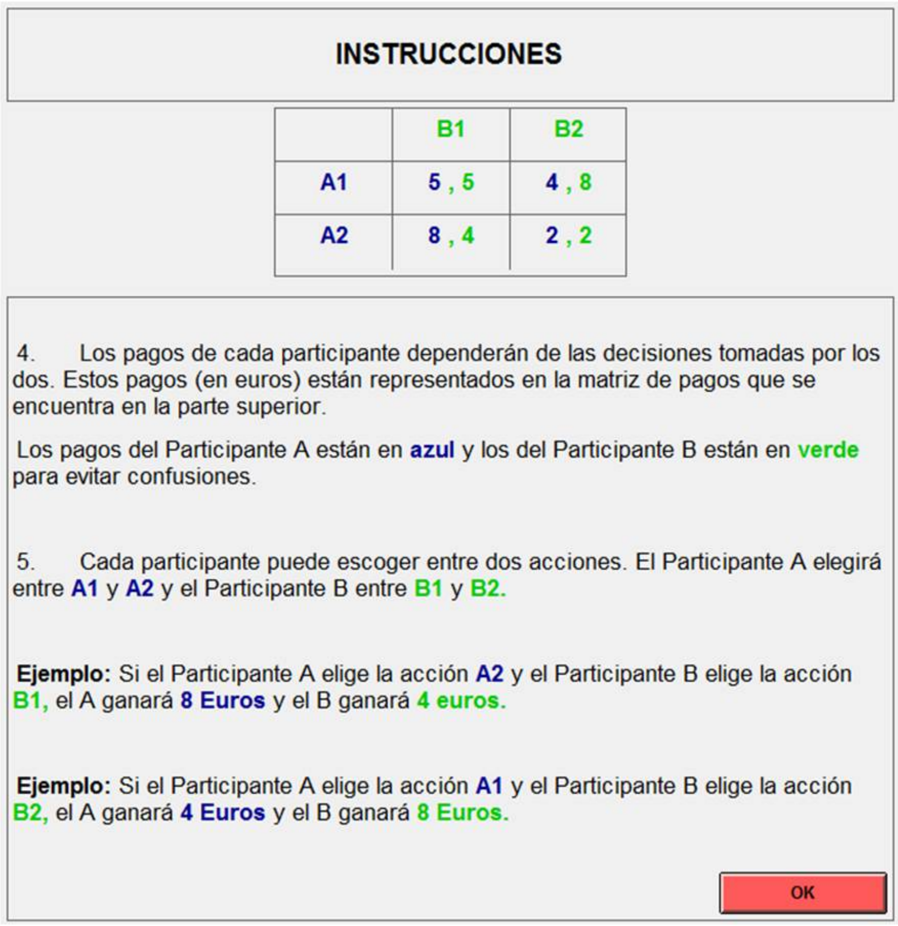


1. To assure the anonymity and the confidentiality a code has been randomly assigned to you (yellow card).

2. There are two types of participants: Participants A and Participants B. From the beginning of the experiment the computer will randomly determine if you are Participant A or Participant B.

3. Every Participant A will be randomly paired with a Participant B. The identity of the participant that is paired with you will be completely anonymous before, during and after the experiment.

Figure 3.8: Screen 3: Common instructions



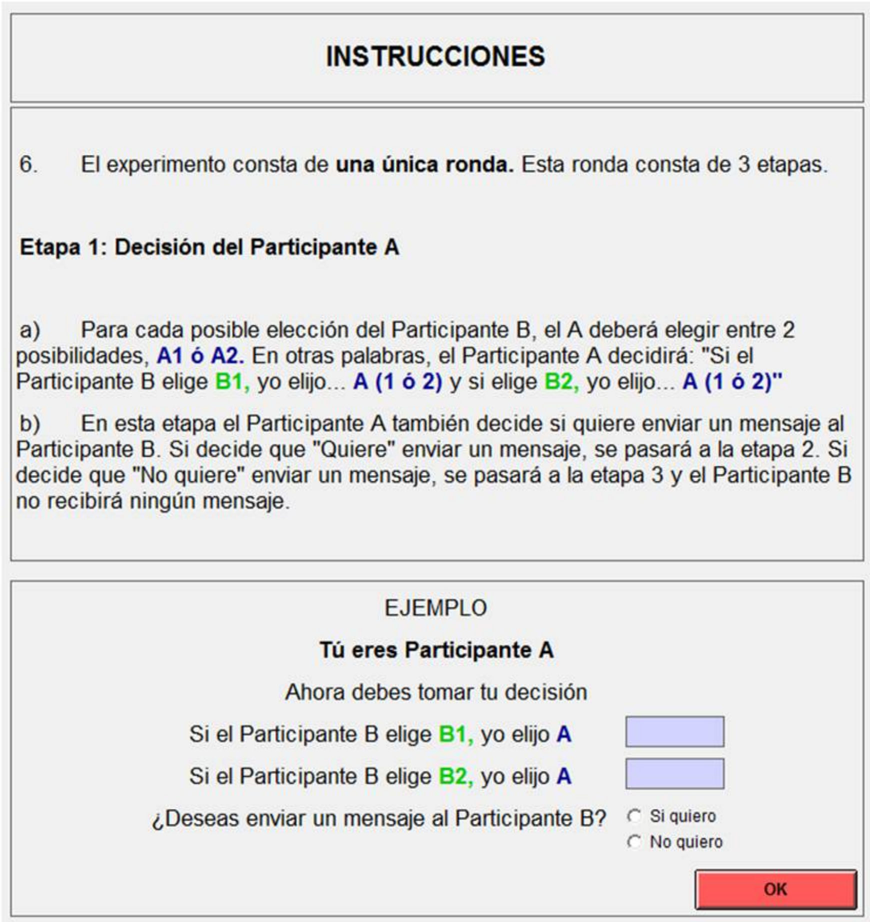
4. The payments of every participant will depend on the decisions made by both. These payments (in euros) are represented in the payoffs matrix that you can find in the top.

5. Every participant can choose among two actions. Participant A will choose among A1 and A2 and Participant B among B1 and B2.

Example: If Participant A chooses the action A2 and Participant B chooses the action B1, A will earn 8 Euros and the B will earn 4 Euros.

Example: If Participant A chooses the action A1 and Participant B chooses the action B2, A will earn 4 Euros and B will earn 8 Euros.

Figure 3.9: Screen 4: NB



6. The experiment consists of only one round. This round has 3 stages.

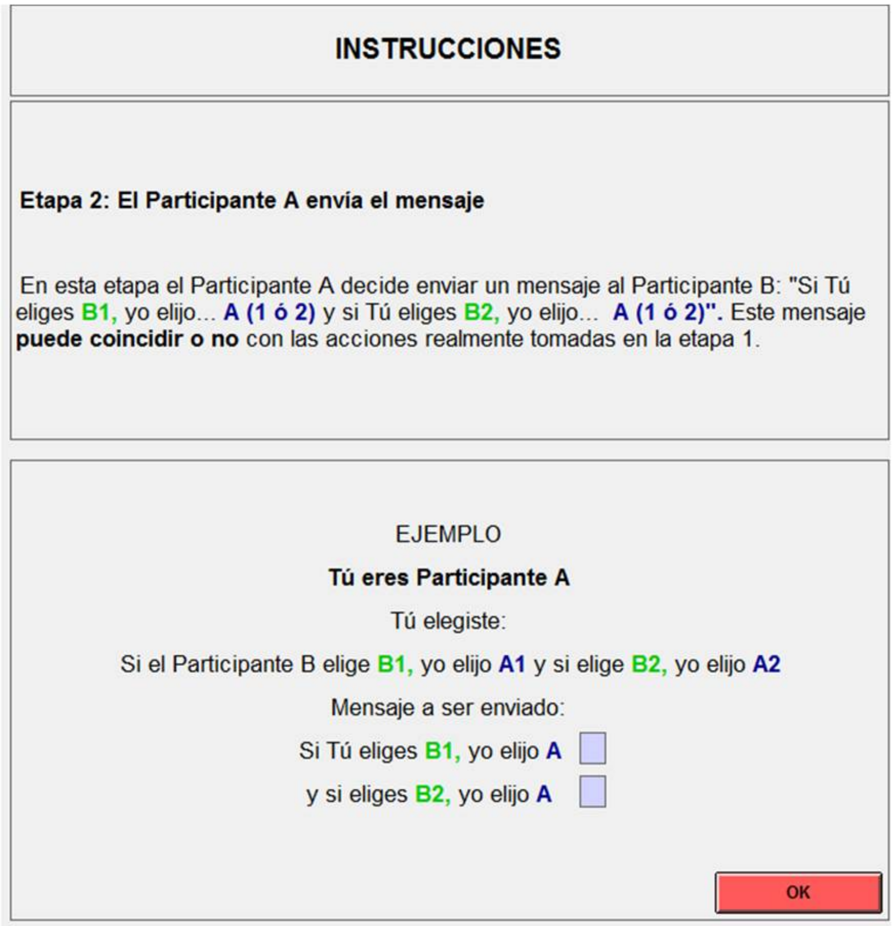
Stage 1: Participant A's decision.

a) For every possible Participant B's choice, A will have to choose among 2 alternatives, A1 or A2. In other words, Participant A will decide: "If the Participant B chooses B1, I choose... A (1 or 2) and if he chooses B2, I choose... A (1 or 2)"

b) In this stage the Participant A also decides if he wants to send a message to the Participant B. If he decides that "he wants" to send a message, he will pass to the stage 2. If he decides that "he does not want" to send a message, he will pass to the stage 3 and the Participant B will not receive any message.

EXAMPLE. You are Participant A.
Now you must take your decision
If Participant B chooses B1, I choose A ( )
If Participant B chooses B2, I choose A ( )
Do you want to send a message to Participant B?

Figure 3.10: Screen 5: NB

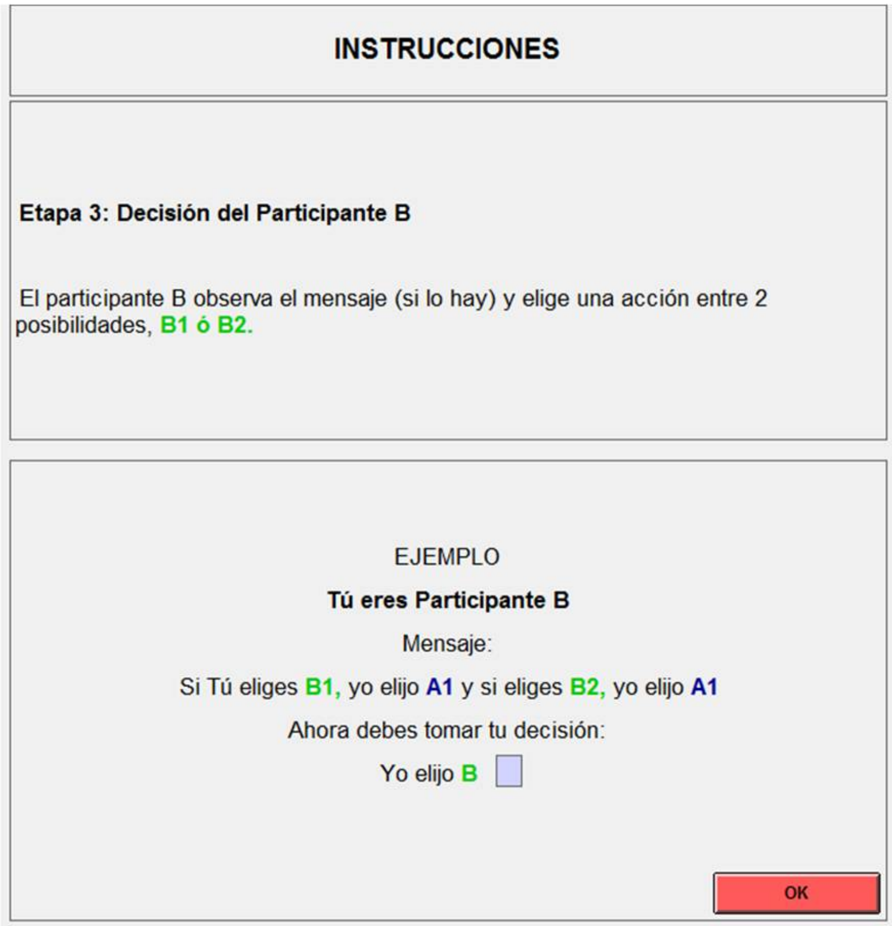


Stage 2: Participant A sends the message.

In this stage Participant A decides to send a message to Participant B: "If You choose B1, I choose... A (1 or 2) and if You choose B2, I choose... A (1 or 2)". This message can coincide or not with the actions really taken in the stage 1.

EXAMPLE. You are Participant A.
You chose: "If You choose B1, I choose A1 and if You choose B2, I choose A2".
Message to being sent:
If You choose B1, I choose A ( )
If You choose B2, I choose A ( )

Figure 3.11: Screen 6: NB

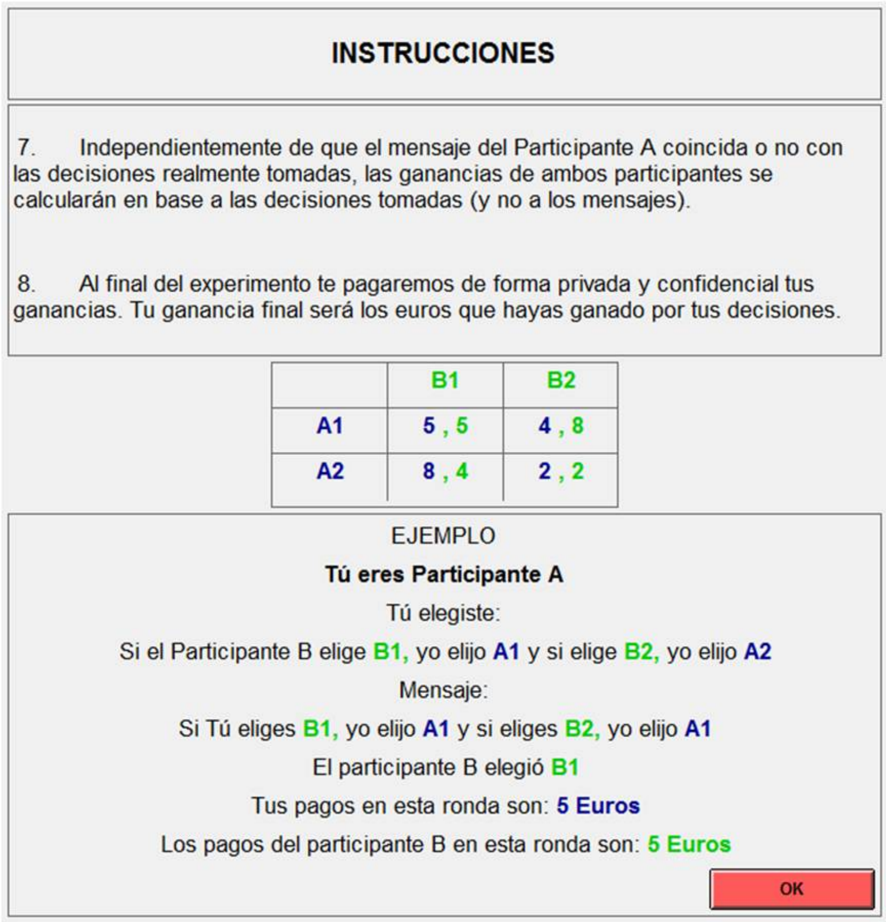


Stage 3: Participant B's decision

Participant B observes the message (if there any) and he chooses an action among 2 alternatives, B1 or B2.

EXAMPLE.
You are Participant B.
Message:
"If You choose B1, I choose A1 and if You choose B2, I choose A1".
Now you must take your decision
I choose B ( )

Figure 3.12: Screen 7: NB



INSTRUCCIONES

7. Independientemente de que el mensaje del Participante A coincida o no con las decisiones realmente tomadas, las ganancias de ambos participantes se calcularán en base a las decisiones tomadas (y no a los mensajes).

8. Al final del experimento te pagaremos de forma privada y confidencial tus ganancias. Tu ganancia final será los euros que hayas ganado por tus decisiones.

	B1	B2
A1	5 , 5	4 , 8
A2	8 , 4	2 , 2

EJEMPLO

Tú eres Participante A

Tú elegiste:

Si el Participante B elige **B1,** yo elijo **A1** y si elige **B2,** yo elijo **A2**

Mensaje:

Si Tú eliges **B1,** yo elijo **A1** y si eliges **B2,** yo elijo **A1**

El participante B eligió **B1**

Tus pagos en esta ronda son: **5 Euros**

Los pagos del participante B en esta ronda son: **5 Euros**

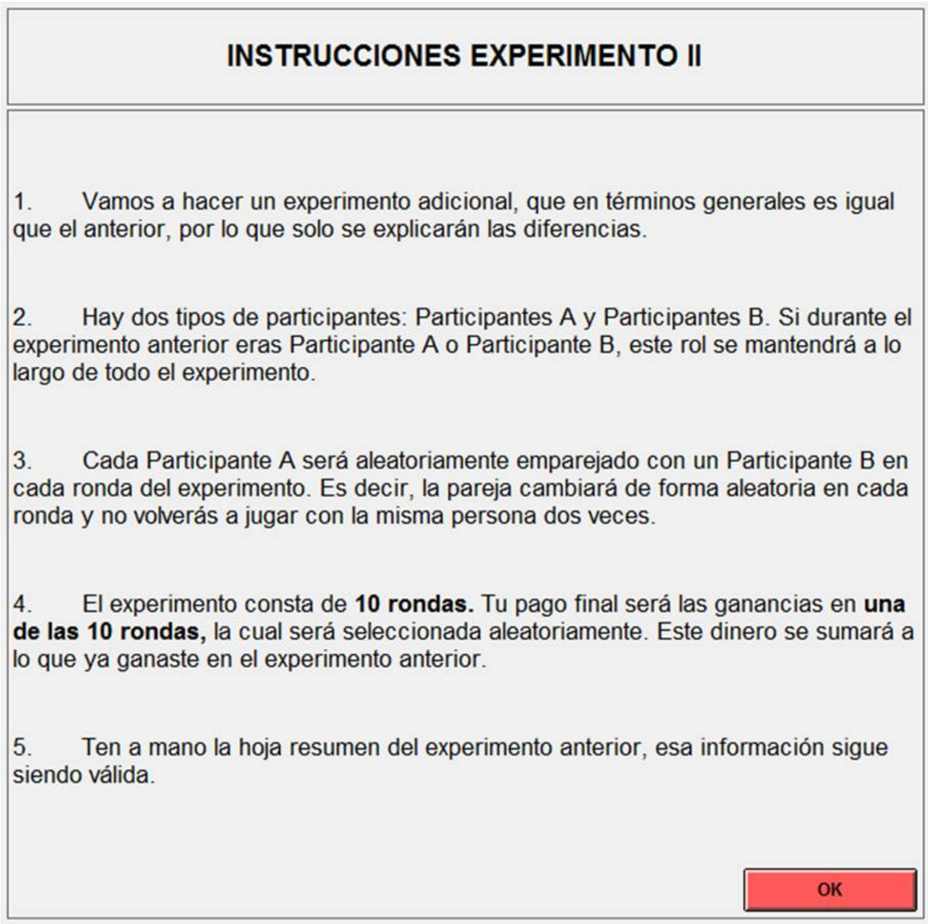
OK

7. Independently of Participant A's message coincides or not with the really taken decisions, the earnings of both participants will be calculated based on decisions (not on messages).

8. At the end of experiment we will pay to you in a private and confidential way. Your final payoff will be the Euros that you have earned according to your decisions.

EXAMPLE. You are Participant A.
You chose: "If You choose B1, I choose A1 and if You choose B2, I choose A2".
Message: "If You choose B1, I choose A1 and if You choose B2, I choose A1".
Participant B choses: B1.
Your payoffs in this round are: 5 Euros.
Participant B's payoffs in this round are: 5 Euros.

Figure 3.13: Surprise restart



EXPERIMENT II

1. We are going to do an additional experiment, which in general terms is like the previous one, therefore only the differences will be explained.

2. There are two types of participants: Participants A and Participants B. If during the previous experiment you are Participant A or Participant B, this role will be kept along the whole experiment.

3. Every Participant A will be randomly paired with a Participant B in every round of the experiment. That is to say, the couple will change randomly in every round and you will not play with the same person two times.

4. The experiment consists of 10 rounds. Your final payment will be the earnings in one of the 10 rounds, which will be randomly selected. This money will be added to what you already have at the previous experiment.

5. Have at hand the summary of the previous experiment, because this information is still valid.

Chapter 4

The Strategic Role of Non-binding Communication

Abstract: This article studies the conditions that may improve bargaining power using threats and promises. We develop a model of strategic communication, based on the *conflict game with perfect information*, in which a noisy *commitment message* is sent by a better-informed sender to a receiver who takes an action that determines the welfare of both. Our model captures different levels of aligned-preferences, for which classical games such as *stag hunt*, *hawk-dove* and *prisoner's dilemma* are particular cases. We characterize the Bayesian Perfect Equilibrium with non-binding messages under *truth-telling beliefs* and *sender's bargaining power* assumptions. Based on our equilibrium selection, we show that the less conflict the game has, the more informative the equilibrium signal is and that less credibility is necessary to implement it.

JEL Code: D03, D74, D82

Keywords: Commitment, threats, promises, credibility, cheap talk.

4.1. Introduction

Bargaining power refers to the relative skills that players have in order to exert influence upon each other. For instance, bargaining power is related to idiosyncratic characteristics. Based on Rubinstein's (1982) dynamic model, a player turns the final outcome onto his favor if he has better outside options or if he is more patient. In addition, following Schelling (1960), bargaining power has also been described as the power to cheat and bluff, the ability to set the best price for oneself and to fool others by pretending this was one's maximum offer. When the union says to the management in a firm, "we will go on strike if you do not meet our demands," or when a nation announces that any military provocation will be responded with nuclear weapons, it is clear that communication has been used for a tactical purpose.

In bargaining theory, *strategic moves* are actions taken prior to playing a subsequent game, with the aim of changing the available strategies, information structure or payoff functions. The aim is to change the opponent's believes, making it credible that the position is unchangeable. Following Selten (1975), the modern formal notion of credibility is subgame perfectness[1]. Nevertheless, we argue that if a message is subgame perfect, then it is neither a threat nor a promise. Let us consider the following examples: a union says to management, "If you increase our salaries, we will be grateful"; or a nation announcement, "under attack, we will defend ourselves". In such cases, credibility is not in doubt, but we could hardly call this a promise or a threat. Schelling (1960) denotes fully credible messages as *warnings*.

Commitment theory was proposed by Schelling (1960)[2], who introduced a tactical approach for communication and credibility inside game theory. Hirshleifer (1987, 2000) and Klein and O'Flaherty (1993) worked on the analysis and characterization of strategic moves in the standard game theory framework. In the same way, Crawford and Sobel (1982) formally showed that an informed agent could reveal his

[1]Schelling developed the notion of credibility as the outcome that survives iterated elimination of weakly dominated strategies. We know that in the context of generic extensive-form games with complete and perfect information, this procedure does indeed work (see Dixit, 2006)

[2]For a general revision of Schelling's contribution to economic theory, see Dixit (2006) and Myerson (2009) surveys.

information in order to induce the uninformed agent to make a specific choice.

There are three principal reasons for modeling pre-play communication: information disclosure (signaling), coordination goals (cheap-talk), and strategic influence (in Schelling's sense). Following Farrell (1993) and Farrell and Rabin (1996), the main problem in modeling non-binding messages is the "babbling" equilibrium, were statements mean nothing. However, they showed that cheap talk can convey information in a general signaling environment, displaying a particular equilibrium in which statements are meaningful. In this line, Rabin (1990) developed *credible message profiles*, looking for a meaningful communication equilibrium in cheap-talk games.

Our article contributes to the strategic communication literature in two ways. First, we propose a particular characterization of *warnings*, *threats* and *promises* in the *conflict game with perfect information*, as mutually exclusive categories. For this aim, we first define a sequential protocol in the 2×2 *conflict game* originally proposed by Baliga and Sjöström (2004). This benchmark game is useful because it is a stylized model that captures different levels of aligned-preferences, where the classical games like *stag hunt*, *hawk-dove* and *prisoner's dilemma* are considered as particular cases.

Second, we model strategic moves with non-binding messages, showing that choosing a particular message and its credibility are related with the level of conflict. In this way, the *conflict game with non-binding messages* captures a bargaining situation where people talk about their intentions, by simply using cheap talk. More precisely, we analyze a game where a second player (the sender) can communicate her action plan to the first mover (the receiver)[3]. In fact, the sender must decide after she observes the receiver's choice, but the commitment message is a pre-play move.

We analyze conceptually the importance of three essential elements of commitment theory: the choice of a response rule, the announcement about future actions, and the credibility of messages. We answer the following questions: *what is the motivation behind threats and promises?* and *can binding messages improve the sender's bargaining power?* In this article, *threats* and *promises* are defined as a second

[3] To avoid confusion and gender bias, the sender will be denoted as "she", and the receiver as "he".

mover self-serving announcement, committing in advance how she will play in all conceivable eventualities, as long as it specifies at least one action that it is not her best reply (see Hirshleifer, 2000; Dixit, 2006). With this definition, we argue that binding messages improve the sender's bargaining power in the *perfect information conflict game*, even when it is clear that by assuming binding messages we avoid the problem of credibility.

The next step is to show that credibility is related to the probability that the sender fulfills the action specified in the non-binding message. For this, we highlight that players share a common language, and the literal meaning must be used to evaluate whether a message is credible or not. Hence, the receiver has to believe in the literal meaning of announcements if and only if it is highly probably to face the truth. Technically, we capture this intuition in two axioms: *truth-telling beliefs* and *the sender's bargaining power*. We ask: *are non-binding messages a mechanism to improve the sender's bargaining power?* and *how much credibility is it necessary for a strategic move to be successful?* In equilibrium, we can prove that non-binding messages will convey private information when the conflict is low. On the other hand, if the conflict is high, there are too strong incentives to lie, and cheap talk becomes meaningless. However, even in the worse situation, the non-binding messages can transmit some meaning in equilibrium if the players focus on the possibility of fulfilling threats and promises.

The article is organized as follows. In section 4.2, the 2×2 *conflict game* is described. In section 4.3 the conditioned messages will be analyzed, highlighting definitions of *threats* and *promises*. Section 4.4 presents the model with non-binding messages, showing the importance of response rules, messages and credibility to improve the sender's bargaining power. Finally, section 4.5 concludes.

4.2. Benchmark Model: The 2×2 Conflict Game

The 2×2 *conflict game* is a non-cooperative symmetric environment with two decision makers, Player 1 and 2. The set of players is $N = \{1, 2\}$[4]. Players must choose

[4]In this level of simplicity, players' identity is not relevant, but since the purpose is to model Schelling's strategic moves, in the followings sections Player 2 is going to be a sender of *commitment messages*.

an action $x_i \in X_i = \{d, h\}$, where d represents being dove (peaceful negotiator) or, h, being hawk (aggressive negotiator). The utility function $U_i(x_1, x_2)$ for Player i is defined by the payoffs matrix in Table 4.1, where rows correspond to the own choice, and columns represent the other player's choice.

Table 4.1: The 2 × 2 conflict game

		j: d	j: h
i	d	0	$-\delta$
	h	$\mu - c$	$-c$

We assume that $\delta, \mu, c > 0$. In this sense, Player j's aggression imposes a cost on Player i, given that for each one it is always better if her opponent chooses d. Notice that δ captures the cost of being caught out when the opponent is aggressive, while μ represents a benefit from being more aggressive than the opponent. Player i has a cost c of making the aggressive action. In addition, we assume that $\mu - c > -\delta$ because it is better to be hawk when the opponent is dove than the opposite, and $c \neq \delta$ in order to avoid irrelevant equilibrium conditions.

Under these assumptions, *the* 2×2 *conflict game* contains four particular cases, and following Hirshleifer (1987), they could be ordered by their degree of conflict or aligned-preferences levels.

1. *Dominant strategy dove (DSD):* If $0 > \mu - c$ and $c > \delta$.

2. *Strategic complements (SC):* If $0 > \mu - c$ and $c < \delta$. This case corresponds to the *Stag hunt* coordination game, which formalizes the idea that distrust can cause aggression and escalates into conflict.

3. *Strategic substitutes (SS):* If $0 < \mu - c$ and $c > \delta$. This case corresponds to the *chicken* or *hawk-dove* game. It is a model of preemption and deterrence, where fear makes a player back down.

4. *Dominant strategy hawk (DSH):* If $0 < \mu - c$ and $c < \delta$. This case corresponds to the *prisoner's dilemma*, where the individual incentives lead to an inefficient outcome.

Based on incentives, let us explain the degree of conflict for the four cases considered, from least to highest. In the DSD game the players' interests are well aligned and there is no coordination problem because the Nash Equilibrium is unique in dominant strategies. Therefore, a rational player will always choose to cooperate, which leads to achieve the Pareto-optimal outcome. In the SC game, mutual cooperation (d, d) is a Nash equilibrium, but it is not unique in pure strategies. Here there is a coordination problem between a Pareto-dominant equilibrium (d, d) and a risk-dominant equilibrium (h, h). Clearly, in such an environment the dove strategy implies a higher risk and it is the best reply when they trust each other in a reciprocal cooperation. For this reason, we say that distrust could trigger an aggression escalation.

Table 4.2: Nash equilibria in the 2×2 conflict game

	(x_1^*, x_2^*)	(U_1^*, U_2^*)	Is it Pareto optimal?
DSD	(d, d)	$(0, 0)$	yes
SC	(d, d)	$(0, 0)$	yes
	(h, h)	$(-c, -c)$	no
SS	(d, h)	$(-\delta, \mu - c)$	yes
	(h, d)	$(\mu - c, -\delta)$	yes
DSH	(h, h)	$(-c, -c)$	no

The SS game corresponds to a more complex environment, with both coordination and distributional concerns. If just one player behaves aggressively, this turns into her favor, and pure bargaining determines who the winner is. In this 2×2 environment there is no criterion of an ex-ante clear prediction about individual behavior. Finally, DSH describes the standard social dilemma in which rational behavior leads to an inefficient allocation. The DSH game corresponds to a negotiation environment where the prediction is that rationality guarantees the disagreement outcome. In the next sections, it is shown that achieving mutual cooperation in DSH is not an easy task.

4.3. Response Rules and Commitment Messages

Schelling (1960) distinguishes between two different types of strategic moves: ordinary commitments and threats. An ordinary commitment is the possibility of playing first, announcing that the decision has already been taken and that it is impossible to be changed, which forces the opponent to make the final choice. On the other hand, threats are second player moves, where he convincing pledges to respond to the opponent's choice in a specified contingent way (see Hirshleifer, 2000).

We consider now the conflict game with a sequential decision making protocol. The idea is to capture a richer set of strategies that allows us to model threats and promises as self-serving messages. In addition, the set of conditioned strategies include the possibility of implementing ordinary commitment, because a simple unconditional message is always available for the sender.

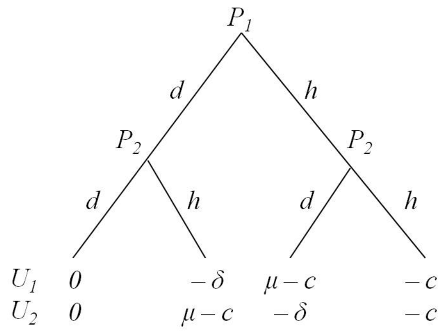
4.3.1. The Conflict Game with Perfect Information

Suppose that Player 1 moves first, and Player 2 observes the action made by Player 1 and make his choice. In theoretical terms, this is a switch from the 2×2 strategic game to the extensive game with perfect information in Figure 4.1. A strategy for Player 2 is a function that assigns an action $x_2 \in \{d, h\}$ to each possible action of Player 1, $x_1 \in \{d, h\}$. Thus, the set of strategies for Player 2 is $X_2^c = \{dd, dh, hd, hh\}$, where $x_2^c = x_2^d x_2^h$ represents a possible reaction rule, such that, the first component x_2^d denotes the action that will be carried out if Player 1 plays d, and the second component x_2^h is the action in case that 1 plays h[5]. The set of strategies for Player 1 is $X_1 = \{d, h\}$.

In this sequential game with perfect information a strategy profile is (x_1, x_2^c). Therefore, the utility function $U_i(x_1, x_2^c)$ is defined by $U_i(d, x_2^d x_2^h) = U_i(d, x_2^d)$ and $U_i(h, x_2^d x_2^h)$ $= U_i(h, x_2^h)$, based in the 2×2 payoff matrix presented before. As the set of strategies profiles becomes wider, the predictions based on the Nash equilibrium are less relevant. Thus, in *the conflict game with perfect information* the applicable equilibrium concept is the Subgame Perfect Nash equilibrium (SPNE).

[5] Note that the subscript specifies the player for each strategy, and the superscript c highlights when the strategy is a conditional move, not a single action.

Figure 4.1: The Conflict game with perfect information



Definition 1 ***SPNE:*** *The strategy profile* (x_1^*, x_2^{c*}) *is a SPNE in the conflict game with perfect information if and only if* $U_2(x_1^*, x_2^{c*}) \geq U_2(x_1, x_2^c)$ *for every* $x_2^c \in X_2^c$ *and for every* $x_1 \in X_1$*; and* $U_1(x_1^*, x_2^{c*}) \geq U_1(x_1, x_2^{c*})$ *for every* $x_1 \in X_1$.

The strategy $x_2^{c*} = x_2^{d*}x_2^{h*}$ represents the best response for Player 2 in every subgame. In the same way, the strategy x_1^* is the best response for Player 1 when Player 2 chooses x_2^{c*}. By definition and using the payoffs assumptions, it is clear that the strategy $x_2^{c*} = x_2^{d*}x_2^{h*}$ is the unique weakly dominant strategy for Player 2, and in consequence, the reason for Player 1 to forecast his counterpart's behavior based on the common knowledge of rationality. The forecast possibility leads to a first mover advantage, as we can see in the Proposition 1.

Proposition 1 *If* (x_1^*, x_2^{c*}) *is a SPNE in the conflict game with perfect information, then* $U_1(x_1^*, x_2^{c*}) \neq -\delta$ *and* $U_2(x_1^*, x_2^{c*}) \neq \mu - c$.

Proof: Suppose that $U_1(x_1^*, x_2^{c*}) = -\delta$ and $U_2(x_1^*, x_2^{c*}) = \mu - c$, then $\mu - c \geq 0$. If $x_2^{c*} = hd$, then $U_1(d, hd) \geq U_1(h, hd)$ and $-\delta \geq \mu - c$, but by assumption $\mu - c > -\delta$ $\Rightarrow\Leftarrow$. If $x_2^{c*} = hh$, then $U_1(d, hh) \geq U_1(h, hh)$ and $-\delta \geq -c$, and at the same time $U_2(h, hh) \geq U_2(h, hd)$. The only compatible case is $c = \delta$, but by assumption $c \neq \delta$ $\Rightarrow\Leftarrow$. Therefore, $U_1(x_1^*, x_2^{c*}) \neq -\delta$ and $U_2(x_1^*, x_2^{c*}) \neq \mu - c$. $\square$

Player 1 is in an advantageous position in this environment, because he always obtains at least as much as his opponent, and as excepted for the DSH game, his outcomes are always the best. Clearly this advantage is related to the opportunity of playing first, which is the idea behind the ordinary commitment. The SPNE for each game is presented in Table 4.3.

Table 4.3: SPNE in the conflict game with perfect information

	(x_1^*, x_2^{c*})	(U_1^*, U_2^*)	Is it Pareto optimal?
DSD	(d, dd)	$(0, 0)$	yes
SC	(d, dh)	$(0, 0)$	yes
SS	(h, hd)	$(\mu - c, -\delta)$	yes
DSH	(h, hh)	$(-c, -c)$	no

We can see that the possibility to play a response rule is not enough to increase Player 2's bargaining power. For this reason, we now consider the case where Player 2 has the possibility to announce the reaction rule she is going to play, before Player 1 makes his decision. Although the signaling game will be analyzed in the next section, we can already answer at this point the following question: Is it possible to increase Player 2's bargaining power by using binding messages?

4.3.2. Threats and Promises as Binding Messages

Following Schelling (1956), the sender's bargaining power increases if she is able to send a message about the action she is going to play, since with premeditation other alternatives have been rejected. For the receiver it must be clear that this is the unique relevant option. This strategic move can be implemented if it is possible to send binding messages about second mover's future actions. With this kind of communication we are going to show that there always exists a message that allows Player 2 to reach an outcome at least as good as the outcome in the SPNE. By notation, $m_2^c \in X_2^c$ is a conditioned message, where $m_2^c = m_2^d m_2^h$. From now on, Player 2 is the sender and Player 1 is the receiver.

Definition 2 ***Commitment message:*** *$m_2^{c*} \in X_2^c$ is a commitment message if and only if $U_2(x_1^{m*}, m_2^{c*}) \geq U_2(x_1^*, x_2^{c*})$, where $U_1(x_1^{m*}, m_2^{c*}) \geq U_1(x_1, m_2^{c*})$ for every $x_1 \in X_1$.*

The idea behind *commitment messages* is that Player 2 wants to achieve an outcome at least as good as the one without communication, given the receiver's best response. This condition only looks for compatibility of incentives, since the receiver also makes his decisions in a rational way. Following closely the formulations discussed in Schelling (1956), Klein and O'Flaherty (1993) and Hirshleifer (2000), we can classify the commitment messages by using three mutually exclusive categories: *warnings*, *threats* and *promises*.

Definition 3 ***Warning, threat and promise:***

1. *The commitment message* $m_2^{c*} \in X_2^c$ *is a warning if and only if* $m_2^{c*} = x_2^{c*}$.
2. *The commitment message* $m_2^{c*} \in X_2^c$ *is a threat if and only if* $U_2(d, m_2^{c*}) = U_2(d, x_2^{c*})$ *and* $U_2(h, m_2^{c*}) < U_2(h, x_2^{c*})$.
3. *The commitment message* $m_2^{c*} \in X_2^c$ *is a promise if and only if* $U_2(d, m_2^{c*}) < U_2(d, x_2^{c*})$.

The purpose of a *warning* commitment is to confirm that the sender will play her best response after every possible action of the receiver. Schelling does not consider *warnings* as strategic moves, but we prefer to use it in this way because the important characteristic of *warnings* is their fully credibility condition. If agents want to avoid miscoordination related to the common knowledge of rationality, they could communicate it and believe it as well. On the contrary, credibility is an inherent problem in *threats* and *promises*. Definitions 3.2 and 3.3 show that at least one action in the message is not the best response after observing the receiver's choice. In *threats*, the sender does not have any incentive to implement the punishment when the receiver plays hawk. In *promises*, the sender does not have any incentive to fulfill the agreement when the receiver plays dove.

The strategic goal in the conflict game is to deter the opponent of choosing hawk, because by assumption $U_i(x_i, d) > U_i(x_i, h)$. This is exactly the purpose of these binding messages, as appears in Proposition 2.

Proposition 2 *If* $\widehat{m}_2^c$ *is a threat or a promise in the conflict game with perfect information, then* $x_1^{\widehat{m}*} = d$.

Proof: Let be $\widehat{m}_2^c$ a threat or promise. Following Definitions 2 y 3, $U_2(x_1^{\widehat{m}*}, \widehat{m}_2^{c*}) \geq U_2(x_1^*, x_2^{c*})$. Suppose that $x_1^{\widehat{m}*} = h$, then there are two possibilities, $\widehat{m}_2^{c*} = x_2^{c*}$ or $U_2(h, \widehat{m}_2^{c*}) \geq U_2(x_1^*, x_2^{c*})$. If $\widehat{m}_2^{c*} = x_2^{c*}$, then by definition $\widehat{m}_2^{c*}$ is neither a threat nor a promise $\Rightarrow\Leftarrow$. If $U_2(h, \widehat{m}_2^{c*}) \geq U_2(x_1^*, x_2^{c*})$, then $x_1^* = d$ or $x_1^* = h$. If $x_1^* = d$, by assumption $U_2(h, \widehat{m}_2^{c*}) < U_2(d, x_2^{c*})$ $\Rightarrow\Leftarrow$. If $x_1^* = h$ and $\widehat{m}_2^c$ is a threat, then $U_2(h, \widehat{m}_2^{c*}) < U_2(x_2^{c*}, x_1^*) \Rightarrow\Leftarrow$. If $x_1^* = h$ and $\widehat{m}_2^c$ is a promise, it must fulfill $U_2(h, \widehat{m}_2^{c*}) = U_2(h, x_2^{c*})$ and $U_2(d, \widehat{m}_2^{c*}) < U_2(d, x_2^{c*})$. The DSD and SC games are not under consideration because $x_1^* = d$ and in SS y DSH games there is no message such that these conditions are true at the same time $\Rightarrow\Leftarrow$. Therefore $x_1^{\widehat{m}*} = d$. $\square$

With these definitions we characterize the properties of the messages in the four stylized games. This is the meaning of Propositions 3.

Proposition 3 *There exists a commitment message m_2^{c*} such that $U_2(x_1^{m*}, m_2^{c*}) > U_2(x_1^*, x_2^{c*})$ if and only if $\mu - c > 0$.*

Proof: Let consider the message $m_2^c = dh$. By Proposition 1 we know that $U_2(x_1^*, x_2^{c*}) \neq \mu - c$, and by assumption $U_1(d, dh) > U_1(h, dh)$, then $m_2^c = dh$ is a commitment message, because $U_2(d, dh) = 0 \geq U_2(x_1^*, x_2^{c*})$. If $U_2(d, dh) > U_2(x_1^*, x_2^{c*})$, then $0 > U_2(x_1^*, x_2^{c*})$, using Proposition 1 again, we conclude that $(x_1^*, x_2^{c*}) = (h, hx_2^{h*})$. As $x_2^{c*} = hx_2^{h*}$, then $U_2(d, hx_2^{h*}) > U_2(d, dx_2^{h*})$, and we have that $\mu - c > 0$.

The proof in the other direction is as follows. Let is consider $\mu - c > 0$, then $x_2^{c*} = hx_2^{h*}$. Using Proposition 1 we know that $U_1(x_1^*, x_2^{c*}) \neq -\delta$, therefore $x_1^* = h$. Now $U_2(x_1^*, x_2^{c*}) < 0$. As we show in the first part, $m_2^c = dh$ is a commitment message such that $U_2(x_1^{m*}, dh) = 0$. Therefore, there exists a commitment message such that $U_2(x_1^{m*}, m_2^{c*}) > U_2(x_1^*, x_2^{c*})$. $\square$

The condition $\mu - c > 0$ is not satisfied in DSD and SC games, where the level of conflict is low. The implication is that mutual cooperation is achieved in equilibrium and this outcome is the highest for both players. The use of messages under these incentives only needs to confirm the sender's rational choice. If Player 2 plays $m^{c*} = x_2^{c*}$, receiver can anticipate this rational behavior, which is completely credible. This is exactly the essence of the Subgame Perfect Nash Equilibrium proposed by Selten (1975).

An essential element of commitments is to determine under what conditions the receiver must take into account the content of a message, given that the communication purpose is to change the rival's expectations. The characteristic of the *warnings* is to choose the weakly dominant strategy, but in the *threats* or *promises* at least one action is not the best response. Proposition 3 shows that in the SS and DSH games the sender's outcome is strictly higher if she can announce that she does not follow the Subgame Perfect strategy. We summarize these findings in Table 4.4.

Table 4.4: Commitment Messages

	Warning	(U_1^*, U_2^*)	Threat	(U_1, U_2)	Promise	(U_1, U_2)
DSD	(d, dd)	$(0, 0)$	(d, dh)	$(0, 0)$		
SC	(d, dh)	$(0, 0)$	(d, dd)	$(0, 0)$		
SS	(h, hd)	$(\mu - c, -\delta)$	(d, hh)	$(-\delta, \mu - c)$	(d, dh)	$(0, 0)$
DSH	(h, hh)	$(-c, -c)$			(d, dh)	$(0, 0)$

Now we inquire about the credibility of these strategic moves, because if the sender is announcing that she is going to play in an opposite way to the game incentives, this message does not change the receiver's beliefs. The message is not enough to increase the bargaining power. It is necessary that the specified action is actually the one that will be played, or at least that the sender believes it. The objective in the next section is to stress the credibility condition. It is clear that binding messages imply a degree of commitment at a 100% level, but this condition is very restrictive, and it is not a useful way to analyze a real bargaining situation. We are going to prove that for a successful strategic move the degree of commitment must be high enough, although it is not necessary to tell the truth with a probability equal to 1.

4.4. The Conflict Game with Non-binding Messages

The credibility problem is related to how probable it is that the message literally coincides with the real actions. The sender announces her way of playing, but it could be a bluff. In other words, the receiver can believe in the message if it is highly probable that the sender is telling the truth. In order to model this problem the game now proceeds as follows. In the first stage *Nature* assigns a type to Player 2 following a probability distribution. The sender's type is her action plan; her way of playing in case of observing each of the possible receiver's action. In the second stage Player 2 observes her "type" and then sends a signal to Player 1. The signal is the disclosure of her plan, and it can be seem as a noisy message, because it is non-binding. In the last stage, Player 1 processes the signal information and chooses an action, which determines the players' payoffs together with the actual type of Player 2.

Following the intuition behind Rabin's (1990) credible message profile, a commitment announcement can be considered credible if it fulfills the following conditions:

1. When the receiver believes the literal meanings of statements, the types sending the messages obtain their best possible payoff, hence those types will send their messages.
2. The statements are truthful "enough". The "enough" comes from the fact that some types might lie to Player 1 by pooling with a commitment message and the receiver knows it. However, the probability of facing a lie is small enough that it does not affect Player 1's optimal response.

The objective of this section is to formalize these ideas using our benchmark *conflict game*. The strategic credibility problem is intrinsically dynamic, and it makes sense if we consider threats and promises as non-binding messages. Bearing these considerations in mind, from now on the messages are used to announce the sender's intentions, but they are *cheap talk*. Clearly, negotiators use to talk, and in most of the cases it is free, but we show that this fact does not imply that cheap talk is meaningless or irrelevant.

4.4.1. The Conflict Signaling Game

Consider a setup in which Player 2 moves first, Player 1 observes a message from Player 2 but not her type. They choose as follows: In the first stage *Nature* assigns a type θ_2^c for Player 2 as a function that assigns an action $x_2 \in \{d, h\}$ to each action $x_1 \in \{d, h\}$. Player 2's type set is $\Theta_2^c = X_s^c = \{dd, dh, hd, hh\}$, where $\theta_2^c = x_2^d x_2^h$. *Nature* chooses the sender's type following a probability distribution, where $p(\theta_2^c) > 0$ is the probability to choose the type θ_2^c, and $\sum_{\theta_2^c \in \Theta_2^c} p(\theta_2^c) = 1$. In a second stage, Player 2 observes her own type and chooses a message $m_2^c \in \Theta_2^c$. At the final stage, Player 1 observes this message and he must choose an action from his set of strategies $X_1 = \{d, h\}$. The most important characteristic of this *conflict game with non-binding messages* is that communication cannot change the final outcome. Though strategies are more complex in this case, the 2×2 payoff matrix in the *conflict game* is always the way to determine the final payoffs.

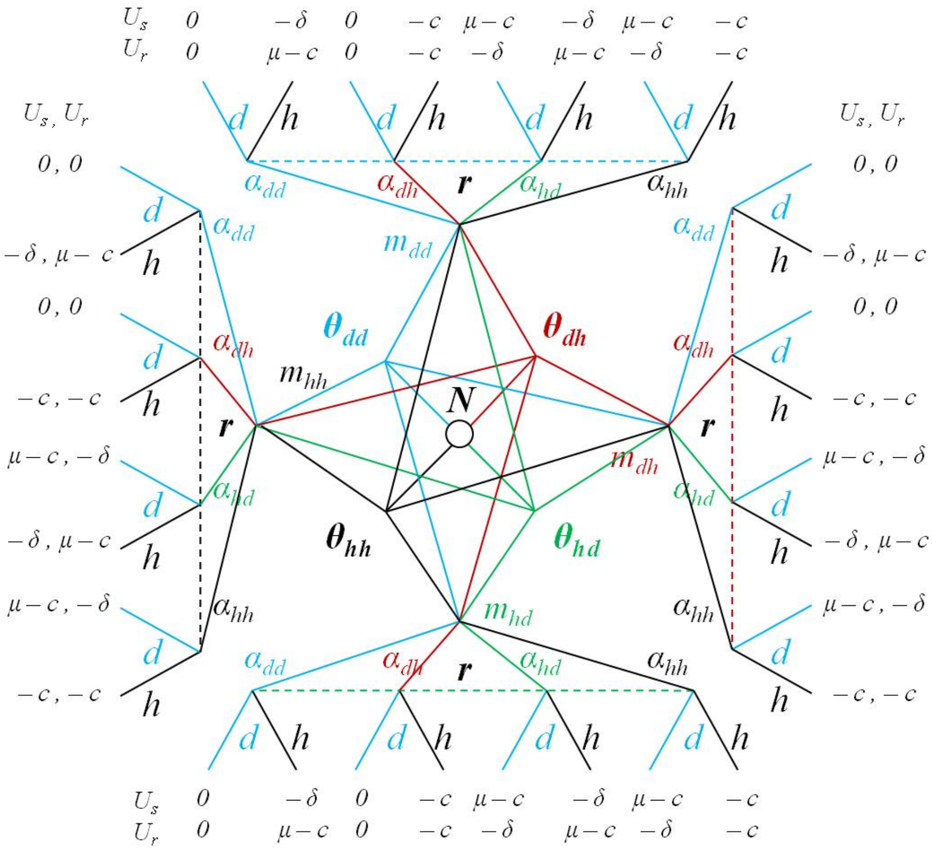
In order to characterize the utility function we need some notation. A message profile $m_2 = (m^c_{dd}, m^c_{dh}, m^c_{hd}, m^c_{hh})$ is a function that assigns a message $m^c_2 \in \Theta^c_2$ to each type $\theta^c_2 \in \Theta^c_2$. The first component $m^c_{dd} \in X^c_2$ is the message chosen in case of observing the type $\theta^c_2 = dd$; the second component $m^c_{dh} \in X^c_2$ is the message chosen in case of observing the type $\theta^c_2 = dh$; and so on. By notation, $m^c_{\theta^c_2} = x^d_2 x^h_2$ is a specific message sent by θ^c_2, and $m_2 = (m^c_{\theta^c_2}, m^c_{-\theta^c_2})$ is a generic message profile with emphasis on the message sent by θ^c_2.

There is imperfect information because the receiver can observe the message, but the sender's type is not observable. Thus, the receiver has four different information sets, depending on the message he faces. A receiver's strategy $x^m_1 = (x^{dd}_1, x^{dh}_1, x^{hd}_1, x^{hh}_1)$ is a function that assigns an action $x_1 \in X_1$, to each message $m^c_2 \in \theta^c_2$, where x^{dd}_1 is the action chosen after observing the message $m^c_2 = dd$, and so on. In addition, $x^m_1 = (x^m_1, x^{-m}_1)$ is a receiver's generic strategy with emphasis in the message he faced. In this case, the superscript m is the way to highlight that the receiver's strategies are a profile of single actions. Therefore, in the *conflict game with non-binding messages* the utility function is: $U_i(x_1^{m^c_{\theta_2}}, x_1^{-m^c_{\theta_2}}, m^c_{\theta^c_2}, m^c_{-\theta^c_2},) = U_i(x_1, x^c_2)$ for $x_1^{m^c_{\theta_2}} = x_1$ and $\theta^c_2 = x^c_2$.

In this specification, messages are payoff irrelevant and what matters is the sender's type. For this reason, it is necessary to define the receiver's beliefs about who is the sender when he observes a specific message. The receiver's belief $\alpha_{\theta^c_2|m^c_2} \geq 0$ is the conditional probability of obtaining the message from a sender of type θ^c_2, given that he observed the message m^c_2. Naturally, $\sum_{\theta^c_2 \in \Theta^c_2} \alpha_{\theta^c_2|m^c_2} = 1$.

All the elements of the *conflict game with non-binding messages* are summarized in Figure 4.2. The most salient characteristics are the four information sets in which the receiver must choose and that the payoffs are independent from the message. For instance, the blue path describes each possible decision for the sender of type *dd*. In the first place, *Nature* chooses the sender's type, in this case $\theta^c_2 = dd$. In the next node, *dd* must choose a message from the 4 possible reaction rules. We say

Figure 4.2: Conflict game with non-binding messages



that dd is telling the truth if she chooses $m^c_{dd} = dd$, leading to the information set at the top. We intentionally plot the game in a star shape in order to highlight the receiver's information sets. At the end, the receiver chooses between d or h, and *cheap talk* implies that there are 4 feasible payoffs.

We recognize that this game has a huge number of Nash Equilibia, consequently, a characterization of this set is not our aim. Our main interest is the communication equilibrium. For this reason the appropriate concept in this case is the Perfect Bayesian Equilibrium.

Definition 4 ***PBE:*** *A Perfect Bayesian Equilibrium is a sender's message profile* $m^*_2 = (m^{c*}_{dd}, m^{c*}_{dh}, m^{c*}_{hd}, m^{c*}_{hh})$*, a receiver's strategy profile* $x^{m*}_1 = (x^{dd*}_1, x^{dh*}_1, x^{hd*}_1, x^{hh*}_1)$ *and a beliefs profile* $\alpha^*_{\theta^c_s|m^c_s}$ *after observing each message* m^c_2*, if the following condi-*

tions are satisfied:

1. m_2^* *is the* $argmax_{m^c_{\theta^c_2} \in \Theta^c_2} U_{\theta^c_2}(x_1^{m*}, m^c_{\theta^c_2}, m^c_{-\theta^c_2})$

2. x_1^{m*} *is the* $argmax_{x_1^{\theta^c_2} \in X_1} \sum_{\theta^c_2 \in \Theta^c_2} \alpha_{\theta^c_2|m^c_2} \cdot U_1(x_1^{m^c_{\theta_2}}, x_1^{-m^c_{\theta_2}}, m_2^*)$

3. $\alpha^*_{\theta^c_2|m^c_2}$ *must be calculated following the bayes' rule based on the message profile* m_2^**. For all* θ^c_2 *who play the message* m_2^{c*}*, the beliefs must be calculated as:* $\alpha_{\theta^c_2|m_2^{c*}} = \frac{p_{\theta^c_2}}{\sum p_{m_2^{c*}}}$

The conditions in this definition are incentive compatibility for each player and Bayesian updating. The first condition requires message $m^{c*}_{\theta^c_2}$ to be optimal for type θ^c_2. The second requires strategy x_1^{m*} to be optimal given the beliefs profile $\alpha^*_{\theta^c_2|m^c_2}$. The last condition is the Bayesian updating, where the receiver's beliefs must be derived via Bayes' rule for each observed message, given the equilibrium message profile m_2^*.

4.4.2. The Commitment Equilibrium Properties

There are, in general, several different equilibria in the *conflict game with non-binding messages*. The objective of this section is to show that a particular equilibrium, that satisfies the following properties, lead to a coordination outcome, given it is both salient and in favor of the sender.

Axiom 1 ***Truth-telling beliefs:*** *If the receiver faces a message* $\widehat{m}_2^{c*} = \widehat{\theta}^c_2$*, then* $\alpha_{\widehat{\theta}^c_2|\widehat{m}^c_2} > 0$*. If the message* $\overline{m}^c_2 = \overline{\theta}^c_2$ *is not part of the messages profile* m_2^**, then* $\alpha_{\overline{\theta}^c_2|\overline{m}^c_2} = 1$.

Axiom 2 ***Senders' bargaining power:*** *If* $m^{c*}_{\theta^c_2}$ *is part of the messages profile* m_2^**, then* $x_1^{m^{c*}_{\theta^c_2}} = d$.

Following Farrell and Rabin (1996) we assume that people in real life do not seem to lie as much, or question each other's statements as much, as the game theoretic predictions state. Axiom 1 captures the intuition that for people it is natural to take seriously the literal meaning of a message. This does not mean that they believe everything they hear. It rather states that they use the meaning as a starting point and then assess credibility, which involves questioning in the form of: "why would she want me to think that? Does she have incentives to actually carry out what she says?"

More precisely, *truth-telling beliefs* emphasize that in equilibrium, when the receiver faces a particular message, its literal meaning is that actually the sender has the intention of playing in this way. Thus, the probability of facing truth-telling messages must be higher than zero. In the same way, when the sender does not choose a particular message, she is signaling that there are no incentives to make the receiver believe this, given that the receiver's best response is h. Therefore, we can assume that the receiver must fully believe in the message, because both players understand that the purpose of the strategic move is to induce the receiver to play d. If the sender is signaling the opposite, she is showing her true type by mistake.

Axiom 2 captures the use of communication as a mean to influence the receiver to play dove. The Nash Equilibrium implies that players must take the other players' strategies as given and then they look for their best respond. Commitment theory implies an additional step, where players recognize that opponents are fully rational. Based on this fact, they evaluate different techniques for turning the other's behavior into their favor. In our case, the sender asks herself: "This is the outcome I would like from this game; is there anything I can do to bring it about?"

In order to characterize a communication equilibrium, we first focus in the completely separator message profile, when the sender is telling the truth. Naturally, $m^c_{\theta^c_2}$ is a truth-telling message if and only if $m^c_{\widehat{\theta}^c_2} = \widehat{\theta}^c_2$. Unfortunately, in Proposition 4 we have a negative result.

Proposition 4 *The completely truth-telling messages profile* $m_2 = (dd, dh, hd, hh)$ *cannot be part of any PBE of the conflict game with non-binding messages.*

Proof: Consider the senders' types $\theta^c_{dh} = dh$ and $\theta^c_{hd} = hd$. If m^*_2 is a completely truth-telling message, then $\alpha^*_{dh|dh} = 1$ and $\alpha^*_{hd|hd} = 1$. By assumption $U_1(d, x_1^{-dh}, dh, m^c_{-dh}) = 0$ and $U_1(h, x_1^{-dh}, dh, m^c_{-dh}) = -c$, then $x_1^{dh*} = d$. In the same way, $U_1(d, x_1^{-hd}, hd, m^c_{-hd}) = -\delta$ and $U_1(h, x_1^{-hd}, hd, m^c_{-hd}) = \mu - c$, then $x_1^{hd*} = h$. Therefore, the utility for the sender is $U_{dh}(d, x_1^{-dh}, dh, m^c_{-dh}) = 0$ and $U_{hd}(d, x_1^{-hd}, hd, m^c_{-hd}) = -\delta$. These conditions imply that the sender type hd has incentives to deviate and $m_2 = (dd, dh, hd, hh)$ cannot be part of any PBE. □

Proposition 5 shows that the completely truth telling message profile is not an equilibrium in the conflict game. The problem lies in the sender type *hd*, because revealing her actual type is not incentive compatible and there always exists at least one successful message to induce the counterpart to play dove. For this reason, we can ask whether there exists some message that induce the sender to reveal her actual type, but at the same time lead to a successful strategic move. Definition 5 is the bridge between *non-binding messages* and *commitment messages* presented in the previous section.

Definition 5 ***Self-committing message:*** *Let be $\widehat{m}^{c*}_{\widehat{\theta}^c_2}$ a truth-telling message and $\alpha_{\widehat{\theta}^c_2|\widehat{m}^{c*}_2(\widehat{\theta}^c_2)} = 1$. $\widehat{m}^{c*}_{\widehat{\theta}^c_2}$ is a self-committing message if and only if $U_{\widehat{\theta}^c_2}(x^{m*}_1, \widehat{m}^{c*}_{\widehat{\theta}^c_2}, \widehat{m}^{c*}_{-\widehat{\theta}^c_2}) \geq U_{\widehat{\theta}^c_2}(x^{m*}_1, m^{c}_{\widehat{\theta}^c_2}, \widehat{m}^{c*}_{-\widehat{\theta}^c_2})$, for every $m^c_{\widehat{\theta}^c_2} \in \Theta^c_2$.*

We introduce the *self-committing* message property because we want to stress that a strategic move is a two-stage process. Talking not only reveals information, but also changes the speaker's incentives through changing what she expects the listener to do. A message is *self-committing*, if believed, it creates incentives for the sender to fulfill it (see Rabin, 1990). The idea behind a *threat* or a *promise* is to implement some risk for the opponent in order to influence him, but this implies a risk for the sender too. This fact has led to associating strategic moves with slightly rational behaviors, when actually in order to be executed, a very detailed evaluation of the consequences is needed. Proposition 5 and it's Corollary explain the relation between the conditioned messages and the incentives to tell the truth.

Proposition 5 *Let be $\widehat{m}^c_2 = \widehat{m}^{c*}_{\widehat{\theta}^c_2}$ a commitment message in the conflict game with perfect information. If $x^{m*(\widehat{\theta}^c_s)}_1 = d$, then $\widehat{m}^{c*}_{\widehat{\theta}^c_2}$ is a self-committing message.*

Proof: Let be $\widehat{m}^c_2 = \widehat{m}^{c*}_{\widehat{\theta}^c_2}$ a commitment message in *the conflict game with perfect information* and $x^{m*(\widehat{\theta}^c_s)}_1 = d$. If $\widehat{m}^{c*}_{\widehat{\theta}^c_2} = \widehat{m}^c_2$ is not a self-committing message, then another message $\overline{m}^c_{\widehat{\theta}^c_2}$ must exist such that $U_{\widehat{\theta}^c_2}(d, x^{-m*(\widehat{\theta}^c_2)}_1, \widehat{m}^{c*}_{\widehat{\theta}^c_2}, m^{c*}_{-\widehat{\theta}^c_2}) < U_{\widehat{\theta}^c_2}(x^{m*}_1, \overline{m}^c_{\widehat{\theta}^c_2}, m^{c*}_{-\widehat{\theta}^c_2})$. Given the payoff assumptions, $U_{\widehat{\theta}^c_2}(d, x^{-m*(\widehat{\theta}^c_2)}_r, \widehat{m}^{c*}_{\widehat{\theta}^c_2}, m^{c*}_{-\widehat{\theta}^c_2}) \geq U_{\widehat{\theta}^c_2}(x^{m*}_1, m^{c*}_{\widehat{\theta}^c_2}, m^{c*}_{-\widehat{\theta}^c_2})$ for every $m^{c*}_{\widehat{\theta}^c_2} \in \Theta^c_2 \Rightarrow\Leftarrow$. Therefore, $\widehat{m}^{c*}_{\widehat{\theta}^c_2} = \widehat{m}^c_2$ is a self-committing message. □

Corollary to Proposition 5: *If $\widehat{m}^c_2$ is a threat or a promise in the conflict game with perfect information, then $\widehat{m}^{c*}_{\widehat{\theta}^c_2} = \widehat{m}^c_2$ is a self-committing message.*

Proof: Using Proposition 2 and Proposition 5.

As we can see in the *conflict game with perfect information*, in the DD and SC games the *warning* is the way to reach the best outcome. If we consider the possibility to send non-binding messages when the sender's type is equal to a *warning* strategy, then revealing her type is self-committing. The problem in the SS and DSH again is more complex given the *warning* message is not self-committing and the way to improve the bargaining power is using a *threat* or a *promise*. This fact leads to a trade-off between choosing a weakly dominant strategy that fails to induce the opponent to play dove, and a strategy that improves her bargaining power but implies a higher risk for both of them.

The required elements for a Bayesian Perfect Equilibrium at each game are shown in Table 4.5 and 4.6. It is important to bear in mind that the beliefs that appear in Table 4.6 are necessary conditions for implementing the EBP presented in Table 4.5, given that they satisfy *truth-telling beliefs* and *sender's bargaining power*.

Table 4.5: Bayesian Perfect Equilibria that satisfy Axioms 1 and 2

	$(m^{c*}_{dd}, m^{c*}_{dh}, m^{c*}_{hd}, m^{c*}_{hh})$	$(x^{dd*}_1, x^{dh*}_1, x^{hd*}_1, x^{hh*}_1)$	$(\alpha^*_{dd\|dd}, \alpha^*_{dh\|dh}, \alpha^*_{hd\|hd}, \alpha^*_{hh\|hh})$
DSD	(dd, dh, dd, hh)	(d, d, h, d)	$(\frac{p_{dd}}{(p_{dd}+p_{hd})}, 1, 1, 1)$
SC	(dd, dh, dd, dh)	(d, d, h, h)	$(\frac{p_{dd}}{(p_{dd}+p_{hd})}, \frac{p_{dh}}{(p_{dh}+p_{hh})}, 1, 1)$
SS	(dh, dh, hh, hh)	(h, d, h, d)	$(1, \frac{p_{dh}}{(p_{dd}+p_{dh})}, 1, \frac{p_{hh}}{(p_{hd}+p_{hh})})$
DSH	(dh, dh, dh, dh)	(h, d, h, h)	$(1, \frac{p_{dh}}{(p_{dd}+p_{dh}+p_{hd}+p_{hh})}, 1, 1)$

The problem of what message must be chosen is as simple as follows in the next algorithm: First the sender tells the truth. If the truth-telling message leads the receiver to play dove, then she does not have any incentive to lie. In the other case, she must find another message to induce the receiver to play dove. If no message leads the receiver to play dove, messages will lack of any purpose, and she will be indifferent between them.

Table 4.5 shows the messages, the receiver's strategies and their belief profiles in a particular equilibrium we argue is the salient. As we can see, in the conflict game

Table 4.6: Beliefs that support the Bayesian Perfect Equilibrium

	Warning	Threat	Promise
DSD	$\alpha^*_{dd\|dd} \geq \frac{\alpha^*_{hd\|dd}(\mu-c+\delta)}{(c-\mu)}$	Truth	
SC	$\alpha^*_{dh\|dh} \geq \frac{\alpha^*_{hh\|dh}(\delta-c)}{c}$	$\alpha^*_{dd\|dd} \geq \frac{\alpha^*_{hd\|dd}(\mu-c+\delta)}{(c-\mu)}$	
SS	Lie	$\alpha^*_{hh\|hh} \geq \frac{\alpha^*_{hd\|hh}(\mu-c+\delta)}{(c-\delta)}$	$\alpha^*_{dh\|dh} \geq \frac{\alpha^*_{dd\|dh}(\mu-c)}{c}$
DSH	Lie		$\alpha^*_{dh\|dh} \geq \frac{\alpha^*_{dd\|dh}(\mu-c)+\alpha^*_{hd\|dh}(\mu-c+\delta)+\alpha^*_{hh\|dh}(\delta-c)}{c}$

the sender is always in favor of those messages where the receiver's best response is dove. In the DSD there are three different messages, in the SC and SS there are two messages and the worst situation is the DSH, where every type of player sends the same message. This fact leads to a first result: if the conflict is high, there are too strong incentives to lie and communication leads to a pooling equilibrium.

In section 4.3 we assume that the sender can communicate a completely credible message in order to influence her counterpart. The question is: how robust is this equilibrium if we reduce the level of commitment? Proposition 6 summarizes the condition for the receiver to choose the dove as the optimal strategy. It is the way for calculating the beliefs that are shown in Table 4.6.

Proposition 6 $x_1^{m(\theta_2^c)*} = d$ *if and only if* $0 \geq (\mu - c)\alpha_{dd|m_2^c} + (-c)\alpha_{dh|m_2^c} + (\mu - c + \delta)\alpha_{hd|m_2^c} + (\delta - c)\alpha_{hh|m_2^c}$.

Proof: We calculate the expected utility for each receiver's strategy and then find the case in which he prefers to choose d.

Based on Proposition 6, the second result is that cheap talk always has meaning in equilibrium. We consider that this equilibrium selection is relevant because the sender focuses on the communication in the literal meanings of the statements, but understands that some level of credibility is necessary to improve her bargaining power. Table 4.6 summarizes the "true enough" property of the statements. Here, the receiver updates his beliefs in a rational way and he chooses to play dove if and only if it is his expected best response. We can interpret the beliefs in Table 4.6 as a threshold, because if this condition is satisfied, the sender is successful in her intention of manipulating the receiver's behavior. Beyond doubt, some level of credibility is necessary, but not at a 100% level.

It is clear that if the conflict is high, the commitment threshold is also higher. In DSD and SC games the sender must commit herself to implement the *warning* strategy, which is a weakly dominant strategy. In SS the strategic movement implies a *threat* or a *promise*; formulating an aggressive statement in order to deter the receiver from behaving aggressively. The worst situation is DSH, where there is only one way to avoid the disagreement point, to implement a *promise*. The promise in this game is a commitment that avoids the possibility of exploiting the opponent, because fear can destroy the agreement of mutual cooperation.

In the scope of this paper, threats are not only punishments and promises are not only rewards. There is a credibility problem because these strategic moves imply a lack of freedom in order to avoid the rational self-serving behavior in a simple one step of thinking. The paradox is that this decision is rational if the sender understands that her move can influence other players' choices, because communication is the way to increase her bargaining power. This implies a second level of thinking, such as a forward induction reasoning.

4.5. Conclusions

In this paper we follow Schelling's tactical approach for the analysis of bargaining. In his Essay on Bargaining, Schelling (1956) analyzes situations where subjects watch and interpret each other's behavior, each one better acting taking into account the expectations that he creates. This analysis shows that an opponent with rational beliefs expects the other to try to disorient him and he will ignore the movements he perceives as stagings especially played to win the game.

The commitment theory is somehow counterintuitive, because it might seem strange that someone wants to limit his options. From a strategic point of view, it is possible to specify theoretically in which cases it is profitable to have less options, even affirming that this choice is carried out in a rational way. A player's lack of freedom has a strategic value because it changes the expectations that others have on the future answers that they could give and this can be used in a self-serving way.

The game presented here is a stylized model that captures different levels of conflict, as a simple parameterization for analyzing the ideas of commitment. If we consider the possibility of sending non-binding messages, in equilibrium we can see that there are strong incentives to tell the truth when the conflict is low, and cheap talk will almost fully transmit private information. On the other hand, when conflict is high, there are too strong incentives to bluff and lie, and commitment must be higher in order to give some meaning to non-binding announcements.

In summary, the equilibrium that satisfies *truth-telling beliefs* and *sender's bargaining power* allows us to show that the less conflict the game has, the more informative the equilibrium signal is and less stronger the commitment it is needed to implement it. Our equilibrium selection is based on the assumption that in reality people do not seem to lie as much, or question each other's statements as much, as rational choice theory predicts. For this reason, the *conflict game with non-binding messages* is a good environment to test different game theoretical hypotheses, because it is simple enough to be implemented in the lab.

As Bolton (1998) suggested, bargaining and dilemma games have been developed in experimental economics as fairly separate literatures. For bargaining, the debate has been centered on the role of fairness and the nature of strategic reasoning. For dilemma games, the debate has involved the relative weights that should be given to strategic reputation building, altruism, and reciprocity. The conflict game and the commitment approach could be the right way to build this bridge, because in the same structure we can study all these elements at the same time. This model provides a simple framework to gather and interpret empirical information. In this way, experiments could indicate which parts of the theory are most useful to predict tha subjects' behavior and at the same time, we can identify behavioral parameters that theory does not reliably determine.

Bibliography

Baliga, S. and Sjöström, T. (2004). Arms races and negotiations. *Review of Economic Studies*, 71(2):351–369.

Binmore, K., Shaked, A., and Sutton, J. (1985). Testing noncooperative bargaining theory: A preliminary study. *The American Economic Review*, 75(5):1178–1180.

Bolton, G. (1991). A comparative model of bargaining: Theory and evidence. *The American Economic Review*, 81(5):1096–1136.

Bolton, G. (1998). Bargaining and dilemma games: From laboratory data towards theoretical synthesis. *Experimental Economics*, 1(3):257–281.

Bolton, G. and Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *The American Economic Review*, 90(1):166–193.

Bornstein, G., Budescu, D., and Zamir, S. (1997). Cooperation in Intergroup, N-person, and Two-person games of chicken. *Journal of Conflict Resolution*, 41(3):384–406.

Brandts, J. and Charness, G. (2000). Hot vs. cold: Sequential responses and preference stability in experimental games. *Experimental Economics*, 2(3):227–238.

Brandts, J. and Charness, G. (2003). Truth or consequences: An experiment. *Management Science*, 49(1):116–130.

Crawford, V. (1998). A survey of experiments on communication via cheap talk. *Journal of Economic Theory*, 78(2):286–298.

Crawford, V. and Sobel, J. (1982). Strategic information transmission. *Econometrica*, 50(6):1431–1451.

Croson, R. (1996). Information in ultimatum games: An experimental study. *Journal of Economic Behavior & Organization*, 30(2):197–212.

Croson, R., Boles, T., and Murnighan, K. (2003). Cheap talk in bargaining experiments: lying and threats in ultimatum games. *Journal of Economic Behavior & Organization*, 51(2):143–159.

Dixit, A. (2006). Thomas Schelling's contributions to game theory. *Scandinavian Journal of Economics*, 108(2):213–229.

Dixit, A. and Nalebuff, B. (1993). *Thinking strategically: The competitive edge in business, politics, and everyday life.* Norton, W. W. & Company, Inc.

Duffy, J. and Feltovich, N. (2002). Do actions speak louder than words? An experimental comparison of observation and cheap talk. *Games and Economic Behavior*, 39(1):1–27.

Elster, J. (1996). Rationality and the emotions. *The Economic Journal*, 106(438):1386–1397.

Elster, J. (1998). Emotions and economic theory. *Journal of economic literature*, 36(1):47–74.

Farrell, J. (1993). Meaning and credibility in cheap talk games. *Games and Economic Behavior*, 5:514–531.

Farrell, J. and Rabin, M. (1996). Cheap talk. *Journal of Economic Perspectives*, 10(1):103–118.

Fehr, E. and Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3):817–868.

Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Fischer, S., Güth, W., Müller, W., and Stiehler, A. (2006). From ultimatum to Nash bargaining: Theory and experimental evidence. *Experimental Economics*, 9(1):17–33.

Frank, R. (1988). *Passions within reason: The strategic role of emotions.* W. W. Norton, New York.

Güth, W., Huck, S., and Rapoport, A. (1998). The limitations of the positional order effect: Can it support silent threats and non-equilibrium behavior? *Journal of Economic Behavior & Organization*, 34(2):313–325.

Güth, W., Müller, W., and Spiegel, Y. (2006). Noisy leadership: An experimental approach. *Games and Economic Behavior*, 57(1):37–62.

Hirshleifer, J. (1987). On the emotions as guarantors of threats and promises. In Dupre, J., editor, *The latest on the best: Essays on evolution and optimality*. MIT Press, Cambridge, MA.

Hirshleifer, J. (2000). Game-theoretical interpretations of commitment. *UCLA Department of Economics in its series UCLA Economics Working Papers with number 799*.

Huck, S. and Muller, W. (2005). Burning money and (pseudo) first-mover advantages: An experimental study on forward induction. *Games and Economic Behavior*, 51(1):109–127.

Kagel, J., Kim, C., and Moser, D. (1996). Fairness in ultimatum games with asymmetric information and asymmetric payoffs. *Games and Economic Behavior*, 13(1):100–110.

Kimbrough, E. and Sheremeta, R. (2010). Make Him an Offer He Can't Refuse: Avoiding Conflicts through Side Payments. *Chapman University, Economic Science Institute. Working Paper with number 10-23*, pages 1–31.

Klein, D. and O'Flaherty, B. (1993). A game-theoretic rendering of promises and threats. *Journal of Economic Behavior & Organization*, 21:295–314.

Mitzkewitz, M. and Nagel, R. (1993). Experimental results on ultimatum games with incomplete information. *International Journal of Game Theory*, 22(2):171–198.

Myerson, R. B. (2009). Learning from Schelling's Strategy of Conflict. *Journal of Economic Literature*, 47(4):1109–1125.

Neugebauer, T., Poulsen, A., and Schram, A. (2008). Fairness and reciprocity in the hawk–dove game. *Journal of Economic Behavior & Organization*, 66(2):243–250.

Norton, E., Wang, H., and Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *The Stata Journal*, 4(2):154–167.

Poulsen, A. and Roos, M. (2010). Do people make strategic commitments? Experimental evidence on strategic information avoidance. *Experimental Economics*, 13(2):206–225.

Poulsen, A. and Tan, J. (2007). Information acquisition in the ultimatum game: An experimental study. *Experimental Economics*, 10(4):391–409.

Rabin, M. (1990). Communication between rational agents. *Journal of Economic Theory*, 51(1):144–170.

Roth, A. (1985). *Toward a focal-point theory of bargaining*. Cambridge University Press, Cambridge, U.K.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, 50(1):97–109.

Schelling, T. (1956). An essay on bargaining. *The American Economic Review*, 46(3):281–306.

Schelling, T. (1960). *The strategy of conflict*. Harvard University Press, Cambridge MA.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4:25–55.

Straub, P. and Murnighan, K. (1995). An experimental investigation of ultimatum games: Information, fairness, expectations, and lowest acceptable offers. *Journal of Economic Behavior & Organization*, 27(3):345–364.