Cognitive Empathy in Conflict Situations*

Florian Gauer† and Christoph Kuzmics‡

September 18, 2017

Abstract

Two individuals are involved in a conflict situation in which preferences are ex-ante uncertain. While they eventually learn their own preferences, they have to pay a small cost if they want to secretely learn their opponent's preferences. We show that there is an interval with upper bound less than one and lower bound greater than zero such that, for sufficiently small positive costs of information acquisition, in any Bayesian Nash equilibrium of the resulting game of incomplete information the probability of acquiring information about the opponent's preferences is within this interval.

Keywords: Incomplete Information, Information Acquisition, Theory of Mind, Conflict, Imperfect Empathy

JEL Codes: C72, C73, D03, D74, D82, D83

*We thank Michael Greinecker, Yuval Heller, Jakub Steiner, and seminar participants at the Universities of Bielefeld, Graz, and Oxford, and at the CERGE-EI in Prague for useful comments and suggestions.

†Center for Mathematical Economics, Bielefeld University, Germany. Email: fgauer86@gmail.com. This research was carried out within the International Research Training Group "Economic Behavior and Interaction Models" (EBIM) financed by the German Research Foundation (DFG) under contract GRK 1134/2. This work was supported by the COLIBRI initiative of the Karl-Franzens University of Graz, Austria.

‡Department of Economics, University of Graz, Austria. Email: christoph.kuzmics@uni-graz.at.

If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle.

— Sun Tzu, The Art of War, approximately 500BC, 1910 translation by Lionel Giles

1 Introduction

We consider situations of strategic interaction in which players know their own cardinal preferences but are ex-ante uncertain about their opponent's cardinal preferences.[1] Players can, however, at some small positive cost, secretely acquire full information about their opponent's cardinal preferences.

The question we pose is this: Will players acquire this information about their opponent's preferences? Let us first survey cases in which this is certainly not, or at least not necessarily, the case.[2] If it is common knowledge that at least one player has a dominant strategy then players derive no benefit from learning anything further about the intensity of each other's preferences and will not acquire this information for any level of cost. If it is common knowledge that players have coordination preferences, then there is certainly an equilibrium in which all individuals play one and the same action regardless of their precise cardinal preferences, in which case players again do not benefit from learning their opponents exact cardinal preferences.[3] But what if the game is one of conflict? That is the game is such that there is no pure strategy profile that both players would agree to be good. Consider the archetypical such game of matching pennies (or one could very similarly think of rock-scissors-paper) with slight payoff disturbances that make the game not necessarily zero-sum, but that preserve the ordinal preferences.

[1] One wonders how both parties succumb in battle, following Sun Tzu's quote, if both know neither their enemy nor themselves.

[2] A trivial case is given by the case in which the cost of acquiring this information is prohibitively high. Then players will certainly not acquire this information. In the paper we, therefore, focus on the case of small costs.

[3] See Section 4.2 for a more detailed discussion of this case.

In any such game with complete information there would only be one Nash equilibrium and this one Nash equilibrium is in completely mixed strategies. These mixed strategies are such that players randomize in such a way as to make each other indifferent. But in order to do so a player would presumably need to know her opponent's payoffs. The refined question we, thus, ask in this paper is: Will players in Bayesian conflict games with incomplete information choose to secretely learn their opponent's preferences?

There is a literature about the value of information that could be helpful here. This literature, starting with Hirshleifer (1971) and including Kamien et al. (1990), Bassan et al. (1997), Gossner and Mertens (2001), Bassan et al. (2003), Lehrer and Rosenberg (2006, 2010), Peski (2008), and Gossner (2010), investigates the value of information under the assumption that whenever a player has additional information, this fact is known by the opponent. In this case having information can be bad for a player.

We are here, however, interested in the case when the possible acquisition of information is secret. The opponent knows that a player could acquire information, but does not observe whether or not she did. More to the point of our question here is Neyman (1991), who showed that any player who is (by chance) given a more precise signal (about some aspect of the game, such as for instance the opponent preferences) must at least weakly benefit. This is true in all games. The intuition behind this result is simple. In such games every player's behavior is independent of the realized signal of her opponents. Thus, given the opponents' behavior is fixed, any player is facing essentially a decision problem and cannot suffer from more precise information.

This brings us back to our question. Is it true that players in conflict situations will always acquire information about their opponent's preferences if this comes at essentially no costs?

The result of Neyman (1991) would suggest this.[4] Another, more technical, result also suggests that this would be the case. It is well known, see e.g., Harsanyi (1973b) and Kohlberg and Mertens (1986), that the Nash equilibrium correspondence in the space of games is upper hemi-continuous. Note

[4]Note, however, that the result of Neyman (1991) does not imply that more information is necessarily strictly better.

that the game in which costs of acquiring information about the opponent's preferences are zero has an equilibrium (provided we deal with finite normal form games) in which all players acquire this information (see e.g., Proposition 1). If this is the only equilibrium of this game, the upper hemi-continuity of the Nash equilibrium correspondence then implies that, if these costs are positive but close to zero, there must be an equilibrium of this game in which players acquire this information, at least with probability close to one.

We find, however, (in Theorem 1) that there is an upper bound less than one and a lower bound greater than zero such that, for sufficiently small costs of information acquisition, all equilibria of, what we call, Bayesian conflict games have a probability of information acquisition that is between the lower and upper bound.

Bayesian conflict games are such that if preferences were common knowledge then every realization of preference pairs would lead to a complete information game that has a unique Nash equilibrium and that unique Nash equilibrium is in completely mixed strategies. Another way to characterize these games is as follows. Players have common knowledge about the ordinal preferences of both players, and these are in some well specifiable sense opposed to each other. The players are uncertain only about their opponent's cardinal preferences, i.e., the difference in intensity of their preferences for the various outcomes.

1.1 An Overview of the Main Results and their Proofs

We actually provide three results of independent interest. First, we show in Proposition 1 that for any positive costs of information acquisition, Bayesian conflict games cannot have equilibria in which both players acquire this information with probability one. This result has a straightforward proof. Suppose it were the case that both players acquire this information with probability one. Then whenever two types realize, players have complete information and, thus, any Nash equilibrium requires appropriate complete mixing in all such cases. But this implies that both players are always completely indifferent between all actions. Then it does not pay to acquire information about the opponent's preferences.

This argument does, however, not imply the main result that the equilibrium probability of getting this information is bounded away from one as costs diminish.

Our second result, Proposition 2, helps to reconcile the apparent contradiction in our main finding (that players' probability of acquiring information is bounded from one as costs tend to zero) and the fact that the Nash equilibrium correspondence in the space of games is upper hemi-continuous. It states that at zero costs the game not only has a Nash equilibrium in which both players acquire information, but also infinitely many in which the probability of acquiring this information is strictly less than one for both players. Any such equilibrium has the following property: If an outside observer can identify all different preference types and then keeps track of the frequency of play of the various actions conditional on any given preference pair, then these probabilities are exactly the Nash equilibrium probabilities of the game in which this preference pair is common knowledge. In other words, in any matching of any two preference types the two types play the Nash equilibrium of the corresponding complete information game despite the fact that they do not have complete mutual knowledge of their preferences! Thus, the players manage to make each other indifferent (after all, this is what is required for completely mixed Nash equilibrium play) without fully knowing their opponents' preferences![5]

The third and main result, Theorem 1, then states that there is an upper bound less than one and a lower bound greater than zero such that, for sufficiently small costs of information acquisition, all equilibria of Bayesian conflict games have a probability of information acquisition that is between the lower and upper bound. Both bounds can be proven by appealing to the upper hemi-continuity of the Bayesian Nash equilibrium correspondence in

[5]We know, from Aumann and Brandenburger (1995), that a sufficient epistemic condition for player's conjectures about opponent play to be a Nash equilibrium in two-player games is the mutual knowledge of the payoff functions, of rationality, and of the conjectures about opponent play. Proposition 2 then implies that even for Bayesian conflict games the condition of mutual knowledge of the payoff functions is not a necessary one, despite the fact that Nash equilibrium play in such games requires precise mixing and this mixing only depends on the opponent's preferences.

the space of games (as the cost parameter c varies). It follows from Proposition 2 that, if costs c are zero, it cannot be an equilibrium to not acquire information about opponent preferences at all (as we assume that each player has at least two distinct types). While the proof needs some care, the main intuition behind this result is this: if nobody acquires information then all types of one player face the same action distribution of the opponent. But then, typically, these two types do not have the same best response actions and would thus play different action distributions themselves. Typically, their opponent would then prefer to know this to adapt her own play accordingly.

To identify the upper bound we need to realize (much as in the proof of Proposition 1) that for small positive costs it cannot be true in any equilibrium that all types are indifferent between all actions. Thus, at least one type of at least one player must be playing an action distribution that puts zero weight on at least one action. But, because of the upper hemi-continuity of the Bayesian Nash equilibrium correspondence, for small costs c, every player type must still be almost indifferent between all actions. But as, by the lower bound, there is a positive probability that a player is informed, and given that she then plays (at least against some types) an action distribution that is far away from the action distribution that makes her opponent type indifferent, she has to balance this action distribution when informed by an appropriate action distribution when not informed and the latter has to receive strictly positive weight. This weight is a function of the distance between the action distribution that puts zero weight on some actions and the by assumption completely mixed action distribution that makes the opponent indifferent. As there is a non-zero distance between these two that does not depend on the cost of information acquisition, the probability of getting informed must be bounded away from one.

1.2 Additional Related Literature

We use the term acquiring "cognitive empathy" instead of the more general term "acquiring information" for two reasons. First, we want to distinguish the specific model of this paper in which information is exclusively about opponent preferences from models in the literature (see below) inspired perhaps by oligopolistic competition in which the uncertainty is about some parameter that affects all players' payoffs. Second, "cognitive empathy" is a term from the psychological literature that fits our model exactly, and is defined in psychology as the process of understanding another person's perspective (see e.g., Davis, 1983), which can be traced back to at least Köhler (1929), Piaget (1932), and Mead (1934). This is in contrast to "affective empathy" which is defined as a person's emotional response to the emotional state of others (see again Davis, 1983) and the two are not necessarily related.[6] The term "empathy" in everyday language typically refers to "affective empathy". We thus emphasize that we are here studying "cognitive empathy".

There is a literature on information acquisition in oligopoly models as in e.g., Li et al. (1987), Hwang (1993), Hauk and Hurkens (2001), Dimitrova and Schlee (2003), and Jansen (2008), where firms can acquire information about the uncertain market demand before engaging in oligopoly competition. Market demand enters all agents' profit functions, whereas in our model the information a player might acquire is exclusively about the opponent's preferences. More general models in which players acquire information about an uncertain parameter affecting all players' preferences are given in Hellwig and Veldkamp (2009), Myatt and Wallace (2012), and Amir and Lazzati (2014), as well as in Persico (2000) and Bergemann et al. (2009) in a mechanism design context.

Moreover, this paper is related to the literature on the evolution of preferences for strategic interaction, initiated by the so-called "indirect evolutionary approach" of Güth and Yaari (1992) and Güth (1995). Individuals who are randomly matched to engage in a given form of strategic interaction are first given a utility function by nature. Nature works on every player separately and aims to maximize this player's material preferences. Players evaluate outcomes of play with the preferences given to them by nature. There are two kinds of results in this literature. Assuming that individu-

[6]Shamay-Tsoory et al. (2009) find that different areas of the human brain are responsible for "cognitive" and "affective" empathy. Rogers et al. (2007) find that people with Asperger syndrome lack "cognitive" but not "affective" empathy.

als (automatically) observe their opponents' preferences, in many settings non-material preferences arise as nature's optimal choice (see e.g., Koçkesen et al., 2000a,b; Heifetz et al., 2007a,b; Dekel et al., 2007; Herold and Kuzmics, 2009). On the other hand, assuming that individuals cannot observe their opponents' preferences, essentially only allows material preferences as nature's optimal choice (see e.g., Ely and Yilankaya, 2001; Ok and Vega-Redondo, 2001). Robson and Samuelson (2010) argued that the potential observability of preferences should also be subject to evolutionary forces.[7] Some work in that direction has recently been begun by Heller and Mohlin (2015a,b).[8] Our model can be seen as tackling the question of the evolution of observability of preferences without modelling the evolution of preferences.

Another such model is given in Robalino and Robson (2012, 2015). In their model, individuals are interacting in ever changing environments. An individual with "theory of mind" is able to use past experiences of opponent play to predict more quickly how her opponent will play. Thus, even if it is somewhat costly, there is a strict benefit from having a "theory of mind." One could argue that the incomplete information (about opponents' preferences) in our model is somewhat akin to the ever changing environment in Robalino and Robson (2015). Our model has no explicit learning. One could perhaps argue it is implicit in our use of Bayesian Nash equilibrium. Our example of a non-conflict game provides a similar result as that in Robalino and Robson (2015) in that any Bayesian Nash equilibrium must exhibit "full" cognitive empathy, i.e., with probability one. When we focus on conflict games alone, we find a starkly contrasting result in that any Bayesian Nash equilibrium must exhibit "partial" cognitive empathy, i.e., the probability of acquiring empathy is bounded from below as well as from above, even when costs of

[7]Similarly (Samuelson, 2001, p. 228) states: "Together, these papers highlight the dependence of indirect evolutionary models on observable preferences, posing a challenge to the indirect evolutionary approach that can be met only by allowing the question of preference observability to be endogenously determined within the model."

[8]The former is a model in which, while individual preferences evolve, so do individuals' abilities to deceive their opponents. The latter asks the question whether cooperation can be a stable outcome of the evolution of preferences in the prisoners' dilemma when players can observe and condition their play on some of their opponent's past actions (in encounters with other people).

acquiring empathy tend to zero.

This paper is also related to Aumann and Maschler (1972), who provide an example of a complete information bimatrix game, due to John Harsanyi, in order to discuss the relative normative appeal of maxmin and Nash equilibrium strategies. The game is a two-player two-action game and not quite zero sum with a unique Nash equilibrium which is in completely mixed strategies. In this game, Nash equilibrium strategies and maxmin strategies differ for both players. Yet the expected payoff to a given player in the Nash equilibrium is the same as the expected payoff that this player can guarantee herself by playing her maxmin strategy. Pruzhansky (2011) provides a large class of complete information bimatrix games that have this feature. If this is the case, would one not, for this class, recommend players to use their maxmin strategies? In our model, in which players have uncertainty about their opponent's preferences, and therefore in some sense greater uncertainty about their opponent's strategy, one might think that the appeal of maxmin strategies is even greater. Yet, in our model there may be a strict benefit from deviating from maxmin strategies, which we show in Gauer and Kuzmics (2016), an earlier working paper version of this paper.

The literature on level-k thinking (see e.g., Stahl and Wilson, 1994, 1995; Nagel, 1995; Ho et al., 1998; Costa-Gomes et al., 2001; Crawford, 2003; Costa-Gomes and Crawford, 2006; Crawford and Iriberri, 2007) typically finds that individuals engaged in game theory experiments do not all reason in the same way as individuals seem to have different "theories of mind". In that sense, our paper can be loosely interpreted as a model to understand why there may be individuals of different levels of strategic thinking.

Solan and Yariv (2004) consider a sequential model of two-player two-action interaction in which one player chooses a (possibly mixed) action first, then a second player can buy, at some cost, information about the first player's (realized) action before finally then also choosing an action herself. The second player can also choose the precision of the information purchased. The structure of the game is common knowledge. In particular the first player is fully aware that she might be spied upon. Thus "spying" in their model is about the opponent's already determined action with complete information regarding payoffs, whereas in our model "spying" (or cognitive empathy as we call it) is about the opponent's preferences and is simultaneous.

Closest is perhaps Mengel (2012), who studies a model in which individuals play many games and ex ante do not know which game they are playing. Individuals can partition the set of games in any way they like, with the understanding that any two games in the same partition element cannot be distinguished. The individual can condition her action only on the partition element. Adopting a partition comes at some cost, called reasoning costs, and finer partitions are more costly than coarser ones. One difference between Mengel (2012) and what we do here is, therefore, that in our model players always learn their own payoff type, while in Mengel (2012) individuals do not necessarily even learn their own payoff type. Another difference is in the choice of solution concept, we study Bayesian Nash equilibria while Mengel (2012) studies asymptotically stable strategy profiles under some evolutionary process. Both these differences are probably only superficial. The real difference between the two papers is the class of games they study within their respective models. Our main results deal with the case of conflict games. Mengel (2012) does not explicitly study this class. Therefore, the nature of our results is also different.[9]

The rest of the paper is organized as follows. Section 2 states the model. Section 3 provides the main results, and Section 4 provides some additional discussion.

[9] The main results in Mengel (2012) are that strict Nash equilibria, while (evolutionarily) stable if the game is commonly known, can be made unstable under learning across games; that weakly dominated strategies, while unstable if the game is commonly known, can be stable under learning across games; and that, if all games have distinct Nash equilibrium supports, learning across games under small reasoning costs leads to individuals holding the finest partition with probability one. Our paper is silent on all these results as our conflict games do not have strict Nash equilibria, do not have weakly dominated strategies, and are such that all (what we call realized type) games are such that their Nash equilibria all have full support. All our results, thus, add to the results in Mengel (2012). One could probably translate our main result into the language of Mengel (2012) as follows. If having the finest partition in the model of Mengel (2012) is essentially the same as acquiring cognitive empathy in our model, then our result, that in conflict games we expect proper mixing between acquiring empathy and not acquiring it, suggests that, in conflict games, learning across games as in Mengel (2012) would lead to individuals properly mixing between different partitions, including the finest as well as the coarsest.

2 The Model

There are two players, $p \in \{B, R\}$, "blue" and "red". Each player p can have one of a finite number n^p of possible (payoff) types $\theta^p \in \Theta^p$. There are commonly known full support probability distributions over types given by $\mu^p : \Theta^p \to (0, 1]$ for both players $p \in \{B, R\}$. The types of the two players are then independently drawn from their respective distribution. Every type of every player has the same finite set of possible actions at her disposal, given by $A = \{a_1, ..., a_m\}$.[10] Let $\Delta(A)$ denote the set of all probability distributions over A, let $\text{int}(\Delta(A))$ denote the set of all such probability distributions over A with full support, and let $\text{bd}(\Delta(A))$ denote the set of all probability distributions that place probability zero on at least one action.[11] Payoffs to player $p \in \{B, R\}$ are then given by the utility function $u^{\theta^p} : A \times A \to \mathbb{R}$, where the first argument depicts the action taken by player p and the second the one taken by her opponent $-p$. Note that different types have different utility functions and that utility functions do only depend on the chosen action pair and not directly on the opponent's type.

Before players learn their own type, i.e., at the complete ex-ante stage, each of them can independently and secretly invest a cost of $c \geq 0$ in order to acquire cognitive empathy. This cost is then simply subtracted from the player's payoff. A player who acquires empathy then, at the interim stage, learns not only her own type but also the type of her opponent. These player types are then called *informed.* Note, however, that an informed type is not able to observe her opponent's choice of empathy acquisition. We further assume that there is only *no empathy* or *full empathy.* When we speak of a player having *partial empathy* we mean that this player randomizes between no and full empathy. A player who does not acquire empathy learns, at the interim stage, only her own type. The corresponding player types are then

[10]In principle, one could consider action sets of different cardinality for both players. The paper, however, focuses on what we call Bayesian conflict games. A crucial feature of Bayesian conflict games is that its "realized type games" (defined below) have a unique equilibrium and that equilibrium is in completely mixed strategies. One can verify that this implies that the two players must have the same number of actions.

[11]A strategy in $\text{int}(\Delta(A))$ could also be called a completely mixed strategy or an interior point of $\Delta(A)$. A strategy in $\text{bd}(\Delta(A))$ could also be called a boundary point of $\Delta(A)$.

called *uninformed*.

A *strategy* of player $p \in \{B, R\}$ is then given by a pair $\left(\rho^p, (\sigma^{\theta^p})_{\theta^p \in \Theta^p}\right)$ where $\rho^p \in [0, 1]$ is the *probability of empathy (or information) acquisition*, and $\sigma^{\theta^p} : \Theta^{-p} \cup \{\emptyset\} \to \Delta(A)$, the *action strategy*, is the (mixed) action to be played by player p of type $\theta^p \in \Theta^p$ against any opponent of known type $\theta^{-p} \in \Theta^{-p}$, when informed, and of unknown type (which is indicated by the player receiving the uninformative "signal" $\emptyset$), when uninformed.

Our solution concept is Bayesian Nash equilibrium. The paper almost exclusively focusses on what we call *Bayesian conflict games.*[12] For any pair of types $\theta^B \in \Theta^B$ and $\theta^R \in \Theta^R$ we define the *realized type game* as the complete information game that would result if it were common knowledge among the two players that they are of exactly these two types. The Bayesian game is then a *Bayesian conflict game* if every possible realized type game has a unique Nash equilibrium and if this Nash equilibrium is in completely mixed strategies.

3 Results

We first show that for positive costs of empathy acquisition there cannot be an equilibrium of a Bayesian conflict game in which both players choose to acquire empathy with probability one.

Proposition 1. *Consider a Bayesian conflict game. If costs of empathy acquisition are positive, then no strategy profile with full empathy, i.e., with $(\rho^B, \rho^R) = (1, 1)$, can be a Bayesian Nash equilibrium. On the contrary, if costs are zero, there is such a full empathy equilibrium.*

Proof of Proposition 1. Suppose a Bayesian conflict game has an equilibrium with $(\rho^B, \rho^R) = (1, 1)$. Then whenever two types $\theta^B \in \Theta^B$ and $\theta^R \in \Theta^R$ meet, it is common knowledge that this is the case and, as this happens with positive probability, they must play a Nash equilibrium of the corresponding realized type game. Any realized type game by definition has a unique

[12]Section 4.1 provides an example of a non-conflict game and Section 4.2 provides a discussion of empathy acquisition in all 2×2 games.

Nash equilibrium and this Nash equilibrium is in completely mixed strategies. Thus, every type of every player is always indifferent between all her pure actions. Hence, when costs are positive, any player would be better off not acquiring empathy, thus saving $c > 0$, and playing any (mixed) action. Arriving at a contradiction, we therefore have the proof for $c > 0$. Observe, however, that this saving opportunity disappears for $c = 0$ meaning that in this case the above strategy profile is indeed an equilibrium of the Bayesian conflict game. $\square$

Proposition 1 leaves open the possibility that, as costs of empathy acquisition tend to zero, the equilibrium probability of empathy acquisition tends to one. To see that this is not true, we turn to the main result of this paper. It establishes that in any equilibrium of a Bayesian conflict game for any of the two players the probability of empathy acquisition is bounded away from one for all sufficiently small positive costs.

In order to state this theorem we require one additional piece of notation. In a Bayesian conflict game, for any player $p \in \{B, R\}$ of any type $\theta^p \in \Theta^p$ denote by $x(\theta^p) \in \Delta(A)$ the mixed action strategy that, if played by the opponent, makes θ^p indifferent between all actions. One could call $x(\theta^p)$ the *indifference inducing mixed action* of type θ^p. From the assumption of a Bayesian conflict game it follows that for each type there is a unique such indifference inducing mixed action. It also follows that $x(\theta^p) \in \mathrm{int}(\Delta(A))$ for all $\theta^p \in \Theta^p$ and $p \in \{B, R\}$. Furthermore we assume that for every player $p \in \{B, R\}$ there are two types $\theta^p, \tilde{\theta}^p \in \Theta^p$ that are distinct, i.e., such that $x(\theta^p) \neq x(\tilde{\theta}^p)$.

As we are able to express the bounds on empathy acquisition under small costs as a function of the parameters of the Bayesian conflict game, a bit more notation is useful. For any action $a \in A$ and any player $p \in \{B, R\}$ let $x_a(\theta^p)$ denote the a-th coordinate of $x(\theta^p)$, i.e., the probability attached to action a in this type's indifference inducing mixed action. Furthermore, let $x_a^{p,\min} := \min_{\theta^p \in \Theta^p} x_a(\theta^p)$ and $x_a^{p,\max} := \max_{\theta^p \in \Theta^p} x_a(\theta^p)$.

Theorem 1. *Consider a Bayesian conflict game. For each cost $c > 0$ of empathy acquisition let ρ_c^p (for any player $p \in \{B, R\}$) denote the probability* *of empathy acquisition in some Bayesian Nash equilibrium of this Bayesian conflict game. Then*

(i) $\liminf_{c \to 0} \rho^p_c \geq \max_{a \in A} \{x_a^{-p,\max} - x_a^{-p,\min}\} > 0$ *and*

(ii) $\limsup_{c \to 0} \rho^p_c \leq \max_{a \in A} \{1 - x_a^{-p,\min}\} < 1.$

In order to prove this theorem we use a result that is of independent interest as well. It characterizes all Bayesian Nash equilibria of the Bayesian conflict game with costs of empathy acquisition $c = 0$.

Proposition 2. *A strategy profile* $\left(\rho^p, (\sigma^{\theta^p})_{\theta^p \in \Theta^p}\right)$ *for both* $p \in \{B, R\}$ *is a Bayesian Nash equilibrium of a Bayesian conflict game with* $c = 0$ *if and only if*

$$\rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p)\sigma^{\theta^p}(\emptyset) = x(\theta^{-p})$$

for all $p \in \{B, R\}, \theta^p \in \Theta^p, \theta^{-p} \in \Theta^{-p}$.

Proof of Proposition 2. The "if" direction is immediate. If the stated condition is satisfied then every player of every type is completely indifferent between all actions, regardless of whether this player has or does not have empathy, and is indifferent between acquiring empathy and not doing so.

To prove the "only if" direction we start by supposing that the strategy profile is a Bayesian Nash equilibrium but does not satisfy the stated condition and then show that this leads to a contradiction. Suppose, therefore, that there is a player p and that there are types θ^p and θ^{-p} such that $\rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p)\sigma^{\theta^p}(\emptyset) \neq x(\theta^{-p})$. This implies that player $-p$ of type θ^{-p} when observing that she faces opponent type θ^p is not indifferent between all actions. Thus, we must have $\sigma^{\theta^{-p}}(\theta^p) \in \mathrm{bd}(\Delta(A))$. We need to distinguish two cases.

Case 1: Suppose $\rho^{-p} = 1$. But then player p of type θ^p, when she knows she is facing type θ^{-p}, is facing action strategy $\sigma^{\theta^{-p}}(\theta^p) \in \mathrm{bd}(\Delta(A))$. This implies that her best response $\sigma^{\theta^p}(\theta^{-p})$ must also be in $\mathrm{bd}(\Delta(A))$ as she cannot be indifferent between all actions. But then $\sigma^{\theta^{-p}}(\theta^p)$ is not best against $\sigma^{\theta^p}(\theta^{-p})$ as the complete information game with types θ^p and θ^{-p}

only has a completely mixed Nash equilibrium by the fact that the game is a Bayesian conflict game. A contradiction.

Case 2: Suppose $\rho^{-p} < 1$. But then, as there is a positive probability that type θ^{-p} is facing type θ^{p} when type θ^{-p} is uninformed she must also in this case be playing a best response to the behavior of type θ^{p}. Otherwise she would acquire empathy with probability one. Thus, we must have that $\rho^{-p}\sigma^{\theta^{-p}}(\theta^{p}) + (1-\rho^{-p})\sigma^{\theta^{-p}}(\emptyset)$ must also be in $\mathrm{bd}(\Delta(A)$ and the same argument as in case 1 applies. $\square$

Two additional lemmas are helpful for the proof of Theorem 1. Their proofs are given in the appendix.

Lemma 1. *Let* $\left(\rho^{p}, (\sigma^{\theta^{p}})_{\theta^{p}\in\Theta^{p}}\right)$ *with* $p \in \{B, R\}$ *be a Bayesian Nash equilibrium of a Bayesian conflict game with* $c = 0$. *Then* $\rho^{p} \geq \max_{a\in A}\{x_a^{-p,\max} - x_a^{-p,\min}\}$.

Lemma 2. *Let* $\left(\rho^{p}, (\sigma^{\theta^{p}})_{\theta^{p}\in\Theta^{p}}\right)$ *with* $p \in \{B, R\}$ *be a Bayesian Nash equilibrium of a Bayesian conflict game with* $c = 0$. *Suppose that there is a* $p \in \{B, R\}$ *and that there are types* θ^{p} *and* θ^{-p} *such that* $\sigma^{\theta^{p}}(\theta^{-p}) \in bd(\Delta(A))$. *Then* $\rho^{p} \leq \max_{a\in A}\{1 - x_a^{-p,\min}\}$.

Proof of Theorem 1.i. By Lemma 1 in any Bayesian Nash equilibrium with $c = 0$ we have $\rho^{p} \geq \max_{a\in A}\{x_a^{-p,\max} - x_a^{-p,\min}\}$. The result then follows from the fact that the Nash equilibrium correspondence in the space of games (here as c varies) is upper hemi-continuous, see e.g., Harsanyi (1973b) or Kohlberg and Mertens (1986). That $\max_{a\in A}\{x_a^{-p,\max} - x_a^{-p,\min}\} > 0$ follows from the assumption that there are at least two distinct types for each player. $\square$

Proof of Theorem 1.ii. Consider an equilibrium of the Bayesian game with c greater than but sufficiently close to zero, in which (by Theorem 1.i) players acquire empathy with positive probability. Then there must be a type θ^{p} of each player p and an opponent type θ^{-p} such that when θ^{p} knows she is facing θ^{-p} she is not indifferent between all actions. Otherwise she could improve her payoff by not acquiring empathy at all. This implies that $\sigma^{\theta^{p}}(\theta^{-p}) \in \mathrm{bd}(\Delta(A))$. By the upper hemi-continuity of the

Nash equilibrium correspondence (as in the proof of Theorem 1.i) and by Lemma 2 we have that ρ^p is less than a number that is arbitrarily close (as c is close to zero) to $\max_{a\in A}\{1 - \min_{\theta^{-p}\in\Theta^{-p}} x_a(\theta^{-p})\}$. The result then follows, as we do not know the identity of the relevant player type θ^{-p}. That $\max_{a\in A}\{1 - \min_{\theta^{-p}\in\Theta^{-p}} x_a(\theta^{-p})\} < 1$ follows from the definition of a Bayesian conflict game. □

4 Discussion

One could give our model at least two interpretations. The first interpretation is that there are indeed two strategic opponents (a soccer player and goalkeeper engaged in a penalty kick, two lawyers, two military generals, etc.) who are involved in a conflict situation and who can acquire information about their opponent's ex-ante unknown preferences. Given this interpretation, we find that in equilibrium these strategic players do not fully acquire information about their opponent's preferences, even if the cost of doing so is vanishingly small.

A second interpretation is that there are many individuals who are often and randomly engaged in pairwise conflict situations and that nature can endow these individuals (each individual separately) with cognitive empathy, i.e., with the ability to understand opponents' preferences, at some positive cost (e.g., by providing an additional brain function). Under the assumption that nature then guides play to an evolutionary stable state, which must be a Bayesian Nash equilibrium of this game, our results imply that nature endows some but not all of her subjects with cognitive empathy, even if the costs of doing so are essentially zero. Similarly, one could appeal to an appropriate version of the purification argument of Harsanyi (1973a), in which there is (additional) uncertainty over individuals' cost of empathy acquisition such that in equilibrium individuals with costs below some equilibrium threshold acquire empathy while those with costs above this threshold do not acquire empathy.

In the following sections we explore the boundaries of our main result. We Section 4.1 provide an example of a non-conflict game with an equilibrium in which both players acquire cognitive empathy with probability one even for moderate costs. In Section 4.2 we extend this analysis to consider all two-player two-action games with ex-ante uncertain preferences and the possibility of cognitive empathy acquisition. We find that, provided costs for this are not prohibitive, that players always acquire ordinal empathy. That is they will learn their opponent's ordinal preferences. In many cases they will then not want to learn their opponent's cardinal preferences, even if this could be done at almost no cost. Finally, coordination games have both kinds of equilibria, some with no empathy acquisition and some with full empathy acquisition. In Section 4.3 we briefly mention other models of cognitive empathy acquisition one could pursue.

4.1 A Non-Conflict Example

In this subsection we provide, as a point of contrast to our main results, a non-conflict example.

Example 1. *Consider a symmetric setup in which both players $p \in \{B, R\}$ can have one of three types $\Theta^B = \Theta^R = \{\theta_1, \theta_2, \theta_3\}$ chosen uniformly (i.e., $\mu^\theta = \frac{1}{3}$ for all $\theta \in \Theta^p$) for the two players. Both players can choose between two actions H and T. Type θ_1 finds action H strictly dominant, type θ_3 finds action T strictly dominant, and type θ_2 has pure coordination preferences. These payoffs, in matrix form, are given in Figure 1.*

$$u^{\theta_1}: \begin{array}{c|c|c|} & H & T \\ \hline H & 1 & 1 \\ \hline T & 0 & 0 \\ \hline \end{array} \qquad u^{\theta_2}: \begin{array}{c|c|c|} & H & T \\ \hline H & 1 & 0 \\ \hline T & 0 & 1 \\ \hline \end{array} \qquad u^{\theta_3}: \begin{array}{c|c|c|} & H & T \\ \hline H & 0 & 0 \\ \hline T & 1 & 1 \\ \hline \end{array}$$

Figure 1: Payoffs of the non-conflict game in Example 1

For costs of empathy acquisition sufficiently low ($c < \frac{1}{9}$) this game has no equilibrium in which a player acquires empathy with probability less than one. Suppose a player (say blue) attaches positive probability to not acquiring empathy. Red makes her choice of action dependent on her own type with dominant action types playing their dominant actions. Now consider the uninformed coordination type of blue. The best she can do is to play a best response to the given (mixed) action of the coordination type of red. W.l.o.g. let this best response action be H. The uninformed coordination type of blue then receives a payoff of zero against the red type with dominant action T. For blue switching to acquiring empathy with probability one and playing T against the T dominant action type of red is then beneficial if $c < \frac{1}{9}$.[13]

4.2 Learning Ordinal before Cardinal Preferences

Consider the following two-player two-action Bayesian game. The action set for both players is $A = \{L, R\}$. Any player $p \in \{B, R\}$ can have one of finitely many types, $\theta^p \in \Theta^p$ with a full support distribution over these types denoted by μ^p. In two-player two-action games (ordinal) preferences must be of one of four kinds: L-dominant strategy preferences, coordination preferences (this player wants to match her opponent's action), mis-coordination preferences (this player wants to mis-match her opponent's action), and R-dominant strategy preferences. Assume that Θ^p, for both players p, includes at least two cardinally distinct preferences of each of these four ordinal preference types.

Assume that each player, before playing the game and before learning her own payoff, has two stages of empathy acquisition possibilities. First, each player can simultaneously invest a small cost $\gamma > 0$. If a player does invest this small amount of utility she learns her opponent's ordinal type fully. That is, she learns which of the four ordinal preference types her opponent has. If she does not invest this small amount γ, she learns nothing about her opponent's preferences. Each player then learns her own ordinal type fully regardless of her investment choice at the first stage.

Second, and only if she has acquired this ordinal empathy, each player simultaneously can invest another small amount $c > 0$ to learn her opponent's

[13] Suppose individuals choose whether or not to acquire empathy after they learn their own type. Then the two dominant action types do not acquire empathy, but for c small enough ($c < \frac{1}{3}$) coordination types acquire empathy with probability one in any equilibrium of this game.

full cardinal preferences. If she does not pay this amount c she learns nothing beyond what she learnt in the first stage. Players then learn their own cardinal preference type and remain ignorant about their opponent's choices of empathy acquisition.

The analysis in the previous subsection makes clear that, provided costs γ are small enough (for a given pair of prior beliefs μ^B, μ^R), both players will acquire ordinal empathy. Each player could turn out to have coordination preferences in which case she would benefit greatly if she could learn whether her opponent is one of the two dominant strategy types.

Given this, will these players invest the small second cost c? This depends on the realized ordinal type match. If any of the two players is known to have a dominant strategy preference type, neither of the two players benefit form learning the exact cardinal preferences of their opponent. This only leaves three cases: they both have coordination preferences, they both have mis-coordination preferences, or one has coordination and the other mis-coordination preferences. The last case is the case dealt with in greater generality in this paper. Let us turn to the other two cases. Note that they are completely analogous to each other.

We can, thus, focus, on the last case, that both players have (and are commonly known to have) coordination preferences. Will players want to learn their opponent's cardinal preferences at small costs c? It depends. This subgame has multiple equilibria. There are two equilibria in which both players simply play the same action regardless of their cardinal preference types. As this is a coordination game, noone has an incentive to deviate from this, and as the opponent's choice of action is independent of her cardinal preference type, it does not pay to invest c to learn the opponent's cardinal preference type. There are, however, other equilibria as well. If at least two types of a player's opponent play distinct actions, then, provided costs c are sufficiently small, this player strictly benefits from learning her opponent's cardinal preference type.[14]

Putting these insights together we note two things. First, for small costs

[14] Among this class of equilibria is also the utilitarian-efficient one in which both players choose the action that maximizes the sum of utilities.

of empathy acquisition, in any equilibrium of the overall Bayesian two-player two-action game, both players acquire cognitive ordinal empathy (i.e., learn their opponent's ordinal preference type). Second, this game has equilibria (even if costs of empathy acquisition are negligible) in which no-one fully acquires cognitive cardinal empathy.

This discussion suggests that cognitive empathy (a pre-requisite for a well-calibrated affective empathy) provides potentially more benefit in coordination problems, problems in which both parties essentially agree what should be done, than in conflict problems, problems in which both parties disagree what should be done.

4.3 Other Models of Empathy Acquisition

Another model one could consider is one in which the empathy acquisition happens after players learn their own types. We conjecture that nothing much changes in the results in such a model.

If we insist in considering a large set of possible situations, we believe our model of full or no empathy acquisition is perhaps too simple. A more appropriate model in this case would be one of "rational (in)attention" as in the decision theoretic models of Sims (2003, 2006); Matêjka and McKay (2012, 2015). Adapting these models to our strategic interaction setting could be done by allowing players to buy signals about their opponent's preferences of any precision with costs increasing in the information content of these signals. Another model would be to allow individuals to acquire multiple signals of whatever precision, one after the other, about their opponent's preferences, before making their final action decision. While we do not think that the main insight of our paper would change in such a model, especially of the latter variety, such a model might nevertheless add substantial additional insights.

A Additional proofs

A.1 Proof of Lemma 1

Proof. By Proposition 2 we must have $\rho^p\sigma^{\theta^p}(\theta^{-p}) + (1-\rho^p)\sigma^{\theta^p}(\emptyset) = x(\theta^{-p})$ for all $p \in \{B, R\}, \theta^p \in \Theta^p, \theta^{-p} \in \Theta^{-p}$.

Let us fix one type θ^p. To simplify notation let us number the types of player $-p$ from 1 to K and let $y^k = \sigma^{\theta^p}(\theta^{-p})$, $z = \sigma^{\theta^p}(\emptyset)$, and $x^k = x(\theta^{-p})$ with different k for different θ^{-p}.

To find a lower bound for ρ^p we need to solve $\min_{(y^k)_{k=1}^K, z\in\Delta(A)} \rho^p$ subject to $\rho^p y^k + (1-\rho^p)z = x^k$ for all $k = 1, ..., K$. Note that each equation $\rho^p y^k + (1-\rho^p)z = x^k$ is really a system of equations, one for each coordinate (i.e., for each action): $\rho^p y_a^k + (1-\rho^p)z_a = x_a^k$. We then need to find this lower bound for all types θ^p and the maximum of all so found lower bounds is the desired lower bound.

To solve this problem we can solve this problem for each coordinate (action) first. That is, we first solve, for any $a \in A$, $\min_{(y_a^k)_{k=1}^K, z_a\in[0,1]} \rho^p$ subject to $\rho^p y_a^k + (1-\rho^p)z_a = x_a^k$ for all $k = 1, ..., K$. We need to consider three cases. Without loss of generality, for the given $a \in A$, let $x_a^1 \leq x_a^2 \leq ... \leq x_a^K$.

Case 1: Suppose $z_a \leq x_a^1$. For the system of equations to hold we need that $y_a^1 \leq y_a^2 \leq ... \leq y_a^K$. The most constraining equation is then $\rho^p y_a^K + (1-\rho^p)z_a = x_a^K$. As $y_a^K \leq 1$ (as $y^K \in \Delta(A)$) we obtain that $\rho^p \geq \frac{x_a^K - z_a}{1-z_a}$. This bound is decreasing in z_a and as z_a is assumed to be less than or equal to x_a^1 the lowest ρ^p consistent with this system of equations is given by $\frac{x_a^K - x_a^1}{1-x_a^1}$.

Case 2: Suppose $z_a \geq x_a^K$. This case is analogous to the previous one and yields $\rho^p \geq \frac{x_a^K - x_a^1}{x_a^K}$.

Case 3: Neither of the above two cases, however, turn out to be the relevant cases for our goal to identify the lowest ρ^p consistent with the system of equations. This is achieved, as we now prove, by letting $x_a^1 \leq z_a \leq x_a^K$. As in case 1 we need that $y_a^1 \leq y_a^2 \leq ... \leq y_a^K$ to satisfy the system of equations. The ρ^p-minimizing choices of y_a^k's are then to either set $y_a^1 = 0$ or to set $y_a^K = 1$. In the first case we obtain that $\rho^p \geq \frac{z_a - x_a^1}{z_a}$, in the second case we obtain $\rho^p \geq \frac{x_a^K - z_a}{1-z_a}$. One bound is increasing and the other decreasing in

z_a. Thus, the ρ^p-minimizing choice of z_a is that which makes both bounds equal. We, thus, need to solve $\frac{z_a - x_a^1}{z_a} = \frac{x_a^K - z_a}{1 - z_a}$, which yields $z_a = \frac{x_a^1}{(1 - x_a^K) + x_a^1}$ and finally, $\rho^p \geq x_a^K - x_a^1$. This bound is lower than those derived in cases 1 and 2, and is, thus the solution to the problem for coordinate (action) a.

As ρ^p has to be chosen so as to satisfy this system of equations for each coordinate (action) a, we obtain the result. □

A.2 Proof of Lemma 2

Proof. By Proposition 2 we must have $\rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p)\sigma^{\theta^p}(\emptyset) = x(\theta^{-p})$ for all $p \in \{B, R\}, \theta^p \in \Theta^p, \theta^{-p} \in \Theta^{-p}$.

Consider type θ^p. To simplify notation, as in the proof of Lemma 1, let us number the types of player $-p$ from 1 to K and let $y^k = \sigma^{\theta^p}(\theta^{-p})$, $z = \sigma^{\theta^p}(\emptyset)$, and $x^k = x(\theta^{-p})$. The assumption of this Lemma can thus be stated as there is an action a and some $j \in \{1, ..., K\}$ such that $y_a^j = 0$. Without loss of generality let $x_a^1 \leq x_a^2 \leq ... \leq x_a^K$.

Using this notation we then must have $\rho^p y_a^k + (1 - \rho^p) z_a = x_a^k$ for all $k = 1, ..., K$. As $y_a^j = 0$ we need $z_a > x_a^j$. For the system of equations to be satisfied we then need to have that $y_a^1 \leq y_a^2 \leq ... \leq y_a^K$. We thus must have that $y_a^1 = 0$. This implies that $\rho^p \geq \frac{z_a - x_a^1}{z_a}$ and as $\frac{z_a - x_a^1}{z_a}$ is increasing in z_a we must have $\rho^p \leq 1 - x_a^1$, which provides the desired result. □

Bibliography

Amir, R. and Lazzati, N. (2014). Endogenous information acquisition in bayesian games with strategic complementarities. Mimeo, University of Iowa.

Aumann, R. and Brandenburger, A. (1995). Epistemic conditions for nash equilibrium. *Econometrica*, 63(5):1161–1180.

Aumann, R. J. and Maschler, M. (1972). Some thoughts on the minimax principle. *Management Science*, 18(5):54–63.

Bassan, B., Gossner, O., Scarsini, M., and Zamir, S. (2003). Positive value of information in games. *International Journal of Game Theory*, 32(1):17–31.

Bassan, B., Scarsini, M., and Zamir, S. (1997). *"I Don't Want to Know!": Can it be Rational?* Hebrew University of Jerusalem, Center for Rationality and Interactive Decision Theory. Discussion paper 158.

Bergemann, D., Shi, X., and Välimäki, J. (2009). Information acquisition in interdependent value auctions. *Journal of the European Economic Association*, 7(1):61–89.

Costa-Gomes, M., Crawford, V. P., and Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 69(5):1193–1235.

Costa-Gomes, M. A. and Crawford, V. P. (2006). Cognition and behavior in two-person guessing games: An experimental study. *American Economic Review*, 96(5):1737–1768.

Crawford, V. P. (2003). Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review*, 93(1):133–149.

Crawford, V. P. and Iriberri, N. (2007). Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions? *Econometrica*, 75(6):1721–1770.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1):113.

Dekel, E., Ely, J. C., and Yilankaya, O. (2007). Evolution of preferences. *The Review of Economic Studies*, 74(3):685–704.

Dimitrova, M. and Schlee, E. E. (2003). Monopoly, competition and information acquisition. *International Journal of Industrial Organization*, 21(10):1623–1642.

Ely, J. C. and Yilankaya, O. (2001). Nash equilibrium and evolution of preferences. *Journal of Economic Theory*, 97(2):255–272.

Gauer, F. and Kuzmics, C. (2016). Cognitive empathy in conflict situations. IMW Working Paper 551.

Gossner, O. (2010). Ability and knowledge. *Games and Economic Behavior*, 69(1):95–106.

Gossner, O. and Mertens, J.-F. (2001). The value of information in zero-sum games. Université Paris-Nanterre and CORE, Université Catholique de Louvain.

Güth, W. (1995). An evolutionary approach to explaining cooperative behavior by reciprocal incentives. *International Journal of Game Theory*, 24(4):323–344.

Güth, W. and Yaari, M. E. (1992). Explaining reciprocal behavior in a simple strategic game. In *Explaining Process and Change: Approaches to Evolutionary Economics*, pages 23–24. University of Michigan Press.

Harsanyi, J. C. (1973a). Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory*, 2(1):1–23.

Harsanyi, J. C. (1973b). Oddness of the number of equilibrium points: a new proof. *International Journal of Game Theory*, 2(1):235–250.

Hauk, E. and Hurkens, S. (2001). Secret information acquisition in cournot markets. *Economic Theory*, 18(3):661–681.

Heifetz, A., Shannon, C., and Spiegel, Y. (2007a). The dynamic evolution of preferences. *Economic Theory*, 32(2):251–286.

Heifetz, A., Shannon, C., and Spiegel, Y. (2007b). What to maximize if you must. *Journal of Economic Theory*, 133(1):31–57.

Heller, Y. and Mohlin, E. (2015a). Coevolution of deception and preferences: Darwin and Nash meet Machiavelli. Available at SSRN 2490370.

Heller, Y. and Mohlin, E. (2015b). Observations on cooperation. Available at SSRN 2558570.

Hellwig, C. and Veldkamp, L. (2009). Knowing what others know: Coordination motives in information acquisition. *The Review of Economic Studies*, 76(1):223–251.

Herold, F. and Kuzmics, C. (2009). Evolutionary stability of discrimination under observability. *Games and Economic Behavior*, 67(2):542–551.

Hirshleifer, J. (1971). The private and social value of information and the reward to inventive activity. *American Economic Review*, 61(4):561–574.

Ho, T.-H., Camerer, C., and Weigelt, K. (1998). Iterated dominance and iterated best response in experimental "p-beauty contests". *American Economic Review*, 88(4):947–969.

Hwang, H.-S. (1993). Optimal information acquisition for heterogenous duopoly firms. *Journal of Economic Theory*, 59(2):385–402.

Jansen, J. (2008). Information acquisition and strategic disclosure in oligopoly. *Journal of Economics & Management Strategy*, 17(1):113–148.

Kamien, M. I., Tauman, Y., and Zamir, S. (1990). On the value of information in a strategic conflict. *Games and Economic Behavior*, 2(2):129–153.

Koçkesen, L., Ok, E. A., and Sethi, R. (2000a). Evolution of interdependent preferences in aggregative games. *Games and Economic Behavior*, 31(2):303–310.

Koçkesen, L., Ok, E. A., and Sethi, R. (2000b). The strategic advantage of negatively interdependent preferences. *Journal of Economic Theory*, 92(2):274–299.

Kohlberg, E. and Mertens, J.-F. (1986). On the strategic stability of equilibria. *Econometrica*, 54:1003–37.

Köhler, W. (1929). *Gestalt Psychology*. Liveright.

Lehrer, E. and Rosenberg, D. (2006). What restrictions do bayesian games impose on the value of information? *Journal of Mathematical Economics*, 42(3):343–357.

Lehrer, E. and Rosenberg, D. (2010). A note on the evaluation of information in zero-sum repeated games. *Journal of Mathematical Economics*, 46(4):393–399.

Li, L., McKelvey, R. D., and Page, T. (1987). Optimal research for cournot oligopolists. *Journal of Economic Theory*, 42(1):140–166.

Matêjka, F. and McKay, A. (2012). Simple market equilibria with rationally inattentive consumers. *American Economic Review, Papers and Proceedings of the 104th Annual Meeting of the American Economic Association*, 102(3):24–29.

Matêjka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298.

Mead, G. H. (1934). *Mind, Self, and Society: From the Standpoint of a Social Behaviorist*. University of Chicago Press.

Mengel, F. (2012). Learning across games. *Games and Economic Behavior*, 74(2):601–619.

Myatt, D. P. and Wallace, C. (2012). Endogenous information acquisition in coordination games. *The Review of Economic Studies*, 79(1):340–374.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review*, 85(5):1313–1326.

Neyman, A. (1991). The positive value of information. *Games and Economic Behavior*, 3(3):350–355.

Ok, E. and Vega-Redondo, F. (2001). On the evolution of individualistic preferences: An incomplete information scenario. *Journal of Economic Theory*, 97:231–54.

Persico, N. (2000). Information acquisition in auctions. *Econometrica*, 68(1):135–148.

Peski, M. (2008). Comparison of information structures in zero-sum games. *Games and Economic Behavior*, 62(2):732–735.

Piaget, J. (1932). *Le Jugement Moral chez l'Enfant*. Presses Universitaires de France.

Pruzhansky, V. (2011). Some interesting properties of maximin strategies. *International Journal of Game Theory*, 40(2):351–365.

Robalino, N. and Robson, A. J. (2012). The economic approach to "theory of mind". *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599):2224–2233.

Robalino, N. and Robson, A. J. (2015). The evolution of strategic sophistication. *American Economic Review*, forthcoming.

Robson, A. J. and Samuelson, L. (2010). The evolutionary foundations of preferences. *Handbook of Social Economics*, 1:221–310.

Rogers, K., Dziobek, I., Hassenstab, J., Wolf, O. T., and Convit, A. (2007). Who cares? Revisiting empathy in Asperger syndrome. *Journal of Autism and Developmental Disorders*, 37(4):709–715.

Samuelson, L. (2001). Introduction to the evolution of preferences. *Journal of Economic Theory*, 97(2):225–230.

Shamay-Tsoory, S. G., Aharon-Peretz, J., and Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, 132(3):617–627.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Sims, C. A. (2006). Rational inattention: Beyond the linear-quadratic case. *American Economic Review*, 96(2):158–163.

Solan, E. and Yariv, L. (2004). Games with espionage. *Games and Economic Behavior*, 47(1):172–199.

Stahl, D. O. and Wilson, P. W. (1994). Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization*, 25(3):309–327.

Stahl, D. O. and Wilson, P. W. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254.