5-2020

Affective Empathy in Non-Cooperative Games

Jorge Vásquez
Smith College, jvasquez@smith.edu

Marek Weretka
University of Wisconsin-Madison

Affective Empathy in Non-cooperative Games[*]

Jorge Vásquez[†] Marek Weretka[‡]

July 12, 2019

Abstract

In this paper, we examine strategic settings in which players have interdependent preferences. Players' utility functions depend not only on the strategy profile being played, but also on the *realized* utilities of other players. Thus, players' realized utilities are interdependent, capturing the psychological phenomena of *affective empathy* and *emotional contagion*. We offer a solution concept for these *empathetic games* and show that the set of equilibria is non-empty and, generically, finite. Motivated by psychological evidence, we then analyze *sympathetic* and *antipathetic* games. In the former, players' utilities increase in others' realized utilities, capturing unconditional friendship; in the latter, the opposite holds, resembling hostility.

Keywords: Non-paternalistic preferences, Interdependent utilities, Affective empathy, Emotional contagion.

JEL Codes: D64, D90, D91.

[*]We thank Sigurt Angenent, Andrés Carvajal, Drew Fudenberg, David Pearce, Daniela Puzzelo, Ariel Roginsky, Ariel Rubinstein, Larry Samuelson, Ricardo Serrano-Padial, Marciano Siniscalchi, Joe Sobel, Lones Smith, Eyal Winter, and three anonymous referees for their insightful comments.

[†]Smith College, Department of Economics, 10 Elm Street, Northampton, MA 01063, USA; E-mail: `jvasquez@smith.edu`; Web: `sites.google.com/site/jovasquez/`

[‡]University of Wisconsin-Madison, Department of Economics, 1180 Observatory Drive, Madison, WI 53706, USA; E-mail: `weretka@wisc.edu`; Web: `www.ssc.wisc.edu/~mweretka/`

1 Introduction

In this paper, we analyze a novel class of games with interdependent utilities by considering players I whose primitive utility functions, say $\mathcal{U} \equiv (U_i)_{i\in I}$, depend on the strategy profile $s \in \mathcal{S}$ being played and also on the realized utilities of others u_{-i}. An *empathetic game* is a structure $\langle I, \mathcal{S}, \mathcal{U} \rangle$. For every strategy profile $s \in \mathcal{S}$, players' *realized utilities* $(u_i)_{i\in I}$ solve the *interdependent utility system* — namely, $u_i(s) = U_i(s, u_{-i}(s))$ for all $i \in I$. This type of utility-interdependence, also known as *non-paternalistic preferences*,[1] has been used to model (pure) altruism in a variety of settings, ranging from the economics of the family (Becker, 1974; Ray, 1987; Bernheim, 1989; Bergstrom, 1997, 1999; Galperti and Strulovici, 2017)[2] to social networks (Bourlès et al., 2017).[3] More recently, Ray and Vohra (2019) consider non-paternalistic preferences to explore strategic settings with "payoff-based externalities," in which a player's payoff depends on her own action and the realized payoffs of other players.

However, important formal features of this framework have been neglected by the literature. Most papers assume that the utility-interdependence either has a linear structure or satisfies a contraction condition, forcing *unique* realized utilities at any strategy profile or unique reduced-form preferences over outcomes.[4] Thus, in all these instances, one can perform equilibrium analysis by applying standard solution concepts to the induced reduced-form game.[5] In more general settings, little is known about what happens if feedback effects, captured by the interdependent utility system $\mathcal{U}$, do not induce unique reduced-form preferences over outcomes. In such cases, standard game-theoretic solution concepts, such as the Nash equilibrium, are inapplicable. In general, it is unclear how to systematically embed non-paternalistic preferences in strategic settings or how to perform equilibrium analysis.

In this paper, we examine empathetic games with *general utility interdependences*. To this end, we first augment the basic game-theoretic framework to encompass settings with general utility systems $\mathcal{U}$. Specifically, our conceptual contribution begins by endowing

[1]As Ray and Vohra (2019) stated, "...we might derive our happiness or hatred directly from the *extent* to which others are enjoying themselves, and not from *how* they are doing so" (p. 1).

[2]For example, in an intertemporal allocation context with multiple generations, each generation may care about its own consumption and the well-being of other generations (Koopmans, 1960; Saez-Marti and Weibull, 2005; Pearce, 2008).

[3]Non-paternalistic preferences have also been used in other contexts, including national savings (Ramsey, 1928; Phelps and Pollak, 1968), public finance (Barro, 1974), economic growth (Bernheim, 1989), and environmental economics (Dasgupta, 2008).

[4]For instance, Pearce (2008) invoked Hawkins-Simon conditions on the system of marginal utilities to ensure the existence and uniqueness of a reduced-form utility profile with intuitive comparative statics. Likewise, Ray and Vohra (2019) focused on *coherent* games, where coherence is a weaker condition (on the interdependent utility system $\mathcal{U}$) that yields a unique and "stable" vector of payoffs at every action profile.

[5]See Bergstrom (1989); Bernheim and Stark (1988); Lindbeck and Weibull (1988); Fels and Zeckhauser (2008); Pearce (2008); Bourlès et al. (2017); Courty and Engineer (2017); Ray and Vohra (2019).

players with *beliefs* regarding other players' realized utilities. Next, we require these beliefs to be *consistent* with the underlying interdependent utility system $\mathcal{U}$. This means that, for any strategy profile s, individual beliefs about others' utilities can be rationalized by a solution for this utility system. Formally, for any strategy profile s, players' realized utilities solve $u_i(s) = U_i(s, u_{-i}(s))$ for all $i \in I$. Finally, an *equilibrium* is a pair containing a strategy profile and beliefs such that the strategies are mutual best responses, given consistent beliefs.

We then offer technical contributions for an arbitrary empathetic game. Because emotional feedback among players can cause emotional synchronization to explode, we first tackle the questions of existence, (generic) finiteness, and robustness. We establish that under mild technical conditions on the utility functions — namely, smoothness and boundedness — an equilibrium exists and, typically, there is a finite number of them (Proposition 1). Thus, our solution concept puts enough discipline on the endogenous variables, ensuing tight predictions. Proving generic finiteness demands novel mathematical arguments and is our primary technical contribution. Finally, we show that, although our belief consistency condition requires players to know which solution of the utility system $\mathcal{U}$ is realized at any profile in and out of equilibrium, relaxing this requirement for the latter has no impact on equilibrium outcomes (Proposition 2). These results provide a fundamental step for building a useful framework to perform an equilibrium analysis of empathetic games in general settings.

Next, we proceed to characterize equilibrium outcomes. In general, the emotional feedback effects among players can lead to *multiple* consistent realized utilities for some strategy profiles. For instance, without changes in behavior, an "emotional contagion" process may lead players to either "happiness" or "misery," depending on the strength of the feedback effects. As a result, computing equilibria directly from the definition is, generally, difficult. We introduce an auxiliary *maxmin utility function* for each player that depends only on strategy profiles. These functions give players their best-response utility assuming they have "pessimistic" beliefs, meaning that whenever their realized utilities take multiple values they believe their lowest utility will be realized. We show that an outcome is an equilibrium if and only if (a) players' realized utilities are at least their maxmin utility level, and (b) realized utilities are constant across pure strategies played with positive chance (Proposition 3).

We then study how pre-existing relationships among individuals affect how they perceive and experience the emotions of others. According to De Waal (2008), empathy can manifest as either sympathy or antipathy, affecting the emotional contagion process. Our framework allows us to capture relationships by specifying how the primitive utility functions $\mathcal{U}$ are affected by others' realized utilities. In *sympathetic* games, the utility function of each player rises in others' utilities, capturing, e.g., unconditional friendship or love. We find that sympathy, such as love, can indeed lead to perverse outcomes, such as misery for all parties involved. Specifically, in sympathetic games, players realized utilities are positively related and prone to take multiple ordered values. Multiplicity, generically, occurs provided players care less about others as they become more happy, i.e., when marginal sympathy is diminishing. A novel source of social inefficiency is prone to emerge here, for even if players choose a strategy profile that potentially maximizes social welfare, their realized utilities may self-reinforce in an inefficient way. By means of example, we show that, because sympathy can lead to misery, a pair of sympathetic players may prefer to remain unmatched to prevent such an outcome. Thus, a successful partnership seems to require an outset mechanism to reduce this self-reinforcing "social anxiety." These insights are consistent with psychological evidence,[6] which indicates that, although love brings happiness, it may also bring misery and anxiety. In fact, individuals attracted to one another appear to be more inclined to experience social anxiety. Altogether, love does not imply happiness, and the old adage that "misery loves company" appears to hold true.

By contrast, in *antipathetic* games players' utilities fall in those of others, resembling, e.g., unconditional antagonism and hostility. In two-player antipathetic games, the emotional contagion process causes realized utilities to be negatively related. Unlike sympathetic games, social inefficiencies here stem from the suboptimal choice of players' strategies. In one example, we see that matching two players who dislike one another can be supported in equilibrium provided that their emotions are neutralized so that no one can benefit from the dissatisfaction of the other, which seems to be in line with anecdotal evidence.

Outline. We organize the rest of the paper as follows. In §2, we provide a novel psychological foundation for non-paternalistic preferences and further motivation for our analysis. Then we set up the model and provide examples in §3 and analyze general empathetic games in §4. Next, we characterize equilibrium outcomes in §5 and study sympathetic and antipathetic games in §6. Section §7 relates our framework to other interdependent utility models, and §8 concludes. Omitted proofs and supplemental material are in the Appendix.

2 Psychological Foundations

The type of preferences that we study in this paper capture the psychological phenomena of affective empathy and emotional contagion among individuals. Let us explain this assertion.

Empathy, a neurological process that is deeply rooted in our brains,[7] is the innate capacity to experience the feelings of others and is an essential building block of social interactions.[8] It

[6] See www.psychologytoday.com/us/blog/happiness-design/201502/will-love-make-you-happy.

[7] According to social psychology, empathy evolved as a proximate reward mechanism for pro-social behavior (e.g., mutual defense) in order to increase the evolutionary fitness of a given group (Batson, 2011).

[8] Human empathy lies at the center of Adam Smith's "theory of moral sentiments" (Smith, 1822).

is associated with mirror neurons that fire when individuals face emotional stimuli, sparking emotional feedback and contagion.[9] Recent research studies have found that empathy is a fundamental driver of altruism (see, e.g., De Waal, 2008).

Social psychology broadly classifies empathy into two types: *cognitive* and *affective*. Cognitive empathy is a neural ability to *rationally* recognize others' intentions, beliefs, desires, and objectives, and is related to the so-called "theory of mind." Cognitive empathy lies implicitly at the heart of game-theoretic models; for example, psychological games (Geanakoplos et al., 1989) and modern theories of reciprocity (Rabin, 1993) build on the idea of *intention-based preferences*. In contrast, affective empathy relates to the automatic transmission and propagation of emotions in response to others' emotions.[10] Recent research suggests that the brain processes cognitive empathy and affective empathy in different ways (Kalbe et al., 2010). Altogether, this evidence indicates that, from an economic perspective, the type of preferences we consider aims to capture different psychological phenomena, compared to intention-based preferences. We provide a detailed discussion of the differences between empathetic games and psychological games in section §7-B.

An important component of affective empathy is *emotional contagion*. This process is "...relatively *automatic, unintentional, uncontrollable*, and largely inaccessible to conversant awareness..." (Hatfield et al., 2014). In other words, affective empathy causes individuals to, e.g., unconsciously synchronize their own emotions with those of others and, thus, *converge emotionally* (Hatfield et al., 1993; Singer et al., 2004). Results in social psychology suggest that emotional convergence may occur very quickly (in less than one second) during face-to-face interactions (Iacoboni, 2009). From an economic viewpoint, it seems that our utilities are automatically and unintentionally affected by others' utilities; thus, using an interdependent utility system to determine the individuals' realized utilities makes sense.

Emotional contagion usually emerges in face-to-face interactions, as human beings are prone to automatically mimic the expressions, vocalizations, postures, and movements of other people with whom they interact (Hatfield et al., 1993). For instance, when someone smiles, one tends to spontaneously smile back; likewise, an angry facial expression may spark an angry expression on another's face (Hawk et al., 2012). Emotional contagion is also important in settings in which individuals have pre-existing relationships. When family members, friends, or foes love, like, hate, or envy one another, their emotions may impact one another in unexpected ways. The relevance of emotional contagion, however, may also extend, indirectly, to social interactions with unknown individuals[11] because the emotions of

[9] Neuroscientific studies have shown that the observation of pain experienced by others and the experience of pain automatically activate similar regions of the brain (Singer et al., 2004; Jackson et al., 2005).

[10] See, e.g., `http://greatergood.berkeley.edu/topic/empathy/definition`.

[11] Recent studies indicate that empathetic responses are also elicited even when scanned subjects do not

one individual, such as happiness and sadness, may propagate to a larger group of individuals.

This psychological phenomena provides us with further motivation for our definition of equilibrium described in the following section. In particular, in this paper we think of realized utilities as the steady states of an emotional contagion process.

3 Games with Affective Empathy

In what follows, for given sets X and Y, recall that Y^X is the set of all functions $f : X \to Y$. Also, for any non-empty set X, the set $\Delta(X)$ denotes the set of probability measures on X.

3.1 An Empathetic Game and Equilibrium Concept

We consider a finite set of "empathetic" players I. Each player $i \in I$ chooses a strategy s_i from a finite set S_i. Let $s = (s_i)_{i\in I}$ be a strategy profile, $\mathcal{S} = \times_{i\in I} S_i$ the set of all strategy profiles, and $\mathcal{S}_{-i} = \times_{j\in I\setminus\{i\}} S_j$ the set of strategy profiles excluding player i. For any player i, a mixed strategy is a probability distribution $\sigma_i \in \Delta(S_i)$, and a mixed profile is $\sigma = (\sigma_i)_{i\in I} \in \Sigma \equiv \times_{i\in I}\Delta(S_i)$. A mixed strategy of others excluding player i is $\sigma_{-i} \in \Sigma_{-i} \equiv \times_{i\in I\setminus\{i\}}\Delta(S_i)$ and $\sigma_{-i}(s_{-i}) \equiv \prod_{j\neq i} \sigma_j(s_j)$. A *utility function* to player i is a map $U_i : \mathcal{S}\times\mathbb{R}^{I\setminus\{i\}} \to \mathbb{R}$, defined over strategy profiles $s \in \mathcal{S}$ and other players' realized utility profiles $u_{-i} \in \mathbb{R}^{I\setminus\{i\}}$. That is, each player's utility function depends not only on the strategy profile being played, but also on his or her beliefs about the others' final utilities at that profile. This formulation allows us to capture the emotional contagion process associated with affective empathy, where players are affected by the perceived well-being of others. For any profile s, an interdependent *utility system* $\mathcal{U}(s,\cdot) : \mathbb{R}^I \to \mathbb{R}^I$ denotes the map $u \mapsto (U_i(s, u_{-i}))_{i\in I}$. An *empathetic game* is a structure $\Gamma \equiv \langle I, \mathcal{S}, \mathcal{U}\rangle$.

Notice that since players' primitive utility functions do not depend exclusively on strategy profiles, standard solution concepts, such as the Nash equilibrium, cannot be applied here. To bypass this problem, we endow players with beliefs about others' final utilities. For any player i, an *empathetic belief* is a function $e_i : \mathcal{S} \to \mathbb{R}^{I\setminus\{i\}}$, where $e_i(s)$ is the realized utility profile that player i believes her co-players would attain if strategy s was played.

How does the actual utility of player i relate to how others' conjecture i's utility? We follow the rational expectations approach and assume that players' conjectures must be consistent with the underlying model structure. Specifically, the perceived utility of player i must coincide with her actual utility, as in, e.g., Bergstrom (1999). It is useful then to define

know the person in pain; see Singer and Fehr (2005).

the set of justifiable utility profiles, or *utility set*, $\mathbf{U}(s) \equiv \{u \in \mathbb{R}^I : u = \mathcal{U}(s,u)\}$,[12] and say that a profile of beliefs $e = (e_i)_{i\in I}$ is *consistent* if, for any profile s, there exists a *utility profile* $u(s) \in \mathbf{U}(s)$ such that for every player $i \in I$, $e_i(s) = u_{-i}(s)$. That is, each player i has correct beliefs about others' utilities, given others' beliefs e_{-i}. So, if profile s is being played and beliefs are consistent, then player i's utility obeys: $u_i(s) = U_i(s, u_{-i}(s))$ for $u(s) \in \mathbf{U}(s)$.

Before defining our equilibrium notion, call $U_i^R(\sigma|e)$ the *reduced-form utility* of player i, given profile σ and empathetic beliefs e, where:

$$U_i^R(\sigma|e) \equiv \sum_{s\in\mathcal{S}} U_i(s, e_i(s))\sigma(s) \tag{1}$$

Definition 1. *A pair* (e^*, σ^*) *is an* equilibrium *if:*

i) Beliefs e^* *are consistent;*

ii) For each player $i \in I$ *and profile* $\sigma_i \in \Sigma_i$*, we have* $U_i^R(\sigma_i, \sigma_{-i}^*|e^*) \leq U_i^R(\sigma^*|e^*)$.

As usual, an equilibrium is *pure* if σ^* is a degenerate probability distribution. Notice that realized utilities depend on a solution of the interdependent utility system $\mathcal{U}$, which is based on realized, not expected, strategies. Also, after any profile of strategies is observed, players have a common equilibrium expectation about which utility profile applies. Condition i) states that in equilibrium each player correctly infers or perceives others' utilities, given their equilibrium beliefs; thus, beliefs cannot be refuted, given the information available to each player, and are consistent with the logic of a self-confirming equilibrium (Fudenberg and Levine, 1993).[13] Condition ii) is standard and asserts that σ^* must be a Nash equilibrium with respect to reduced-form utilities, given consistent empathetic beliefs e^*.[14]

To close this section, we discuss our belief consistency condition from a psychological perspective. As discussed in §2, we can imagine that belief consistency emerges from an emotional contagion process. This feedback process naturally arises in face-to-face interactions, wherein players "experience" the utility of others through their mimicry or body language without changing their behavior. This process can be formally described with an intuitive tatonnement process. Indeed, consider two players, i and j (to ease notational burden). We venture that player i revises her beliefs upwards if her actual experience of

[12] In §4, we offer sufficient conditions to ensure that the utility set $\mathbf{U}(s)$ is non-empty for every profile s.

[13] In §4, we consider a weaker notion of consistency that allows players to disagree on the utility profile that would be attained at non-equilibrium strategies.

[14] Notice that players have empathetic beliefs over realized — not over expected — utility profiles. So when a mixed strategy is played, players' empathetic beliefs do not depend on how players choose to randomize. Thus, our equilibrium is not an equilibrium of the mixed extension, as is the case in standard games.

player j's utility is greater than her current belief. That is,

$$\dot{e}_i^t(s) = U_j(s, e_j^t(s)) - e_i^t(s), \tag{2}$$

for all players i, j and time $t \geq 0$. Clearly, empathetic beliefs are at rest if and only if they satisfy condition ii). In other words, our consistency condition selects a steady state for process (2). Notice that this justification of belief consistency also applies to mixed strategies as long as the emotional contagion process occurs at an *ex post* stage — namely, after a strategy profile is realized. Finally, in face-to-face interactions, the tatonnement process (2) combined with experimentation with respect to strategies should allow players to, eventually, learn their payoffs in order to assess unilateral deviations.

3.2 Two Examples

A. The First-date Game. Anne (a) and Bob (b) simultaneously choose whether to go on (G) or cancel (C) their first date. If both cancel, their utility is equal to $U_i\left(\left(C, C\right), u_j\right) = 1$. If one of the players goes, the "stood up" party, say i, gets utility $U_i\left(\left(C, G\right), u_j\right) = -1$, which can be interpreted as an ego penalty, while the "canceling" player j obtains $U_j\left(\left(C, G\right), u_j\right) = 1$. Finally, if both players go the date, their face-to-face interaction results in interdependent utilities with i's utility being $U_i\left((G, G), u_j\right) = \sqrt{2u_j}$. This interdependency captures the idea that players' happiness level depends on their perceptions of how happy their partners are and vice-versa. In other words, utilities are interdependent when players choose G and are independent otherwise (as in these cases there is no face-to-face interaction). We summarize this game below:

	G	C
G	$\sqrt{2u_b}, \sqrt{2u_a}$	$-1, 1$
C	$1, -1$	$1, 1$

What does belief consistency mean in this context? When Anne and Bob choose G, they form respective beliefs u_b and u_a about the other's final utility. If these beliefs are consistent, they satisfy $u_a = \sqrt{2u_b}$ and $u_b = \sqrt{2u_a}$. Thus, in principle, strategy profile (G, G) is consistent with two *starkly* distinct outcomes. In one outcome, both players may end up with low utility $(0, 0)$ while in the other they may get $(2, 2)$, namely, $\mathbf{U}(G, G) = \{(0, 0), (2, 2)\}$. As mentioned in §1, the emotional contagion process causes players' utilities to be positively or negatively reinforced. In particular, Anne derives high utility from the date whenever Bob derives high utility from it (and vice-versa). This added multiplicity is important to assess unilateral deviations. If Anne chose G, then Bob's comparison of G and C would be ambiguous, as his utility necessarily depends on how optimistic both players are. In fact, going on a date, i.e., strategy (G, G), with empathetic beliefs $e_i^*(G,G)=2$ for every player i is a pure equilibrium. However, going on a date with pessimistic self-fulfilling beliefs $e_i^*(G,G)=0$ is not an equilibrium, because in this case cancelling C is a profitable unilateral deviation. Thus, Definition 1 tells us exactly which strategy profile and pair of consistent beliefs we shall expect from equilibrium play. Finally, notice that (C, C) is an equilibrium for any beliefs e^*, whereas $\sigma_a^* = \sigma_b^* = (2/3, 1/3)$, where 2/3 is the chance of playing G, with $e_i^*(G,G) = 2$ for $i \in \{a, b\}$ is the *unique,* full-support, mixed equilibrium — for beliefs are consistent, and (σ_a, σ_b) is a mixed Nash equilibrium given empathetic beliefs.

Altogether, the *multiplicity* of consistent beliefs in the previous example shows us that positive attitudes are necessary if one wants two sympathizing people to match. From a social viewpoint, matching may be desirable, because agents could engage in productive behavior, exploiting potential complementarities in their skills. Nevertheless, agents may be reluctant to match, as they recognize that compassion for another can lead them to misery. This fear or "social anxiety" may push agents to stay isolated,[15] which is an undesirable equilibrium configuration from a social perspective. Thus, positive mindsets or optimistic beliefs may drive agents not only to be more productive, but also happier.

B. A Gift-giving Game. Consider a gift-giving game in which player 1 decides whether to make a monetary transfer to player 2. For simplicity, assume $S_1 = \{0, 1\}$, namely, player 1 can either transfer one unit ($s_1 = 1$) or nothing ($s_1 = 0$). Transferring one unit costs player 1 a delivery fee $\phi \in (0, 1)$. Player 2 is passive in that his payoffs depend on player 1's strategy.

First consider a standard game in which players are incentivized to share wealth through a warm-glow effect (Andreoni, 1989). Specifically, player 1 derives no utility from his own consumption, whereas player 2 cares about his and player 1's final income. We capture this setting with utility functions $U_1(s_1, s_2, u_2) = \alpha s_1$ and $U_2(s_1, s_2, u_1) = (\alpha - \phi)s_1$, where $\alpha \in (0, 1)$ controls the marginal utility of money to player 2.

In the unique equilibrium, player 1 chooses a full transfer $s_1^* = 1$ and obtains $u_1^* = \alpha$, whereas player 2 gets $u_2^* = \alpha - \phi$. This result holds regardless of how much player 2 values player's one transfer. Indeed, when the transfer fee is high enough $\phi > \alpha$, player 2 ends up unhappy with negative utility. Although from player 1's perspective she is being altruistic in making the transfer, she does not internalize how her action impacts player 2's final utility.

Let us now introduce affective empathy into this game. For this end, suppose that player 1 cares also about player 2's welfare such that his utility function is $U_1(s_1, s_2, u_2) = \alpha s_1 + \beta u_2$ with $\beta \in (0, 1)$. In this scenario, there is a unique pair of consistent beliefs

[15] Interestingly, one of the most common anxieties for some people involves relationships with peers to whom they are attracted; see `https://en.wikipedia.org/wiki/Social_anxiety`.

$e_1^*(s_1, s_2) = (\alpha - \phi)s_1$ and $e_2^*(s_1, s_2) = \alpha s_1 + \beta(\alpha - \phi)s_1$. This empathetic game also has a unique equilibrium. *Player 1 chooses a full transfer* $s_1^* = 1$ *if and only if the fee is low enough* $\phi < \alpha(1 + \beta)/\beta \equiv \bar{\phi}$. While player 1 does not care *directly* about the fee, she cares about player 2's welfare which in turn depends on the fee ϕ. Indeed, we see that a full transfer $s_1 = 1$ makes both players worse off, provided the fee is high enough $\phi > \bar{\phi}$. Player 1 now internalizes how her behavior impacts the final utility of player 2, and, therefore, her final utility. This simple example shows how interdependent utilities can impact economic behavior in transfer games with empathetic players.[16] We provide more discussion about interdependent utilities and reduced-form preferences in the literature review in §7.

4 Existence, Generic Finiteness, and Other Results

Unlike in games with independent payoffs, here a strategy profile $\sigma \in \Sigma$ does not provide a complete description of what utilities players might expect in an empathetic game. The reason is that one strategy profile might be associated with more than one solution. In Example 3.2, if Anne and Bob choose $\sigma_i(G) = 1/2$, then each player might obtain 1/4, or 3/4, depending on whether payoffs are either low or high at $s = (G, G)$. This ambiguity vanishes once we attach a realized utility to each strategy profile. A *feasible outcome* of a game is a tuple $o \equiv (\sigma, v) \in \Sigma \times \mathbb{R}^I$, where $v = (v_i)_{i \in I}$ is a utility profile with $v_i \equiv \sum_{s \in \mathcal{S}} \prod_j \sigma_j(s_j) u_i(s)$ and $u(s) \in \mathbf{U}(s)$ for every profile s. For any game Γ, call $\mathcal{O}$ the set of feasible outcomes, and $\mathcal{O}^*$ the set of equilibrium outcomes. Because every equilibrium (e^*, σ^*) induces a unique outcome $o^* \in \mathcal{O}^*$, WLOG we focus on equilibrium predictions regarding strategies and realized utilities.

In general, a finite empathetic game may not have an equilibrium when utility functions are unbounded. For example, consider a two-player empathetic game, where for some strategy profile $s \in \mathcal{S}$ utilities are $U_i(s, u_{-i}) = u_{-i} + 1$ for $i = 1, 2$. The emotional contagion causes an "explosion," yielding an empty utility set $\mathbf{U}(s) = \{\emptyset\}$. It follows then that there does not exist a consistent profile of beliefs that could potentially support an equilibrium. Also, similar to standard finite normal-form games, empathetic games can have an infinite number of equilibria. Here, indeterminacy may occur for other reasons. For instance, take a two-player empathetic game with payoffs $U_i(s, u_{-i}) = u_{-i}$ and for $i = 1, 2$ and $s \in \mathcal{S}$. Thus, for every profile s, the utility correspondence coincides with the 45° degree line, namely, $\mathbf{U}(s) = \{u \in \mathbb{R}^2 : u_1 = u_2\}$. Consistency of beliefs implies that $e_i(s) = e_j(s)$ for $i = 1, 2$ and so in equilibrium, anything goes. To avoid these uninteresting cases, we henceforth make the following assumption on the utility system, $\mathcal{U} = (U_i)_{i \in I}$, unless explicitly stated.

[16] Bourlès et al. (2017) provide an analysis of how transfers are shaped by altruistic social networks.

Assumption 1. *For all $i \in I$, $U_i(s, \cdot)$ is continuously differentiable and bounded for all s.*

We use the following definition of genericity. Fix an empathetic game Γ and consider a family of "perturbed" empathetic games $\{\Gamma_p : p \in \mathcal{P}\}$, where $\mathcal{P}$ is a subset of $\mathbb{R}^{I \times \mathcal{S}}$. For any $p \in \mathcal{P}$, the game Γ_p is constructed by perturbing the utility system $\mathcal{U}$ in Γ, so that $U_i^p(s, u_{-i}) \equiv U_i(s, u_{-i}) + p_{i,s}$ for all $i \in I$ and $s \in \mathcal{S}$. A property is *generic*, if for any empathetic game Γ and any open set of perturbations $\mathcal{P}$, there exists a subset of $\bar{\mathcal{P}} \subset \mathcal{P}$ with full Lebesque measure such that the property holds in the game Γ_p for all $p \in \bar{\mathcal{P}}$.

Next, we show that under Assumption 1, our equilibrium notion is well-defined. Assumption 1; in particular, uniform boundedness rules out "explosions" of utilities due to payoff-based feedback effects. Thus, for any strategy profile, a solution to the interdependent utility system exists, and so does a mixed empathetic equilibrium. The smoothness requirement in Assumption 1 is technical, as explained in Remark 1 below.

Proposition 1. *The equilibrium set $\mathcal{O}^*$ is non-empty and generically finite.*

In our equilibrium notion (Definition 1), players understand how their utilities depend on one another and, further, they anticipate correctly the utilities of others when a given strategy profile is played. This holds for on and off path strategies. However, one may wonder what happens if players hold beliefs that are individually justifiable but not jointly consistent. That is, what happens if players misperceived the realized utility of others when assessing a unilateral deviation? Intuitively, off equilibrium, players may not interact face-to-face and the emotional contagion process may not take place. Thus, rational players may hold beliefs that are not consistent as long as they are individually justifiable. To address this issue, we now introduce a weaker belief consistency condition (e.g., no face-to-face interactions off equilibrium) that relies on common knowledge of the interdependent utility system $\mathcal{U}$.

Fix an equilibrium profile $\sigma^* \in \Sigma$. We say that a profile of beliefs e^* is *weakly consistent* if for each $s \in \text{supp}\,(\sigma^*)$ there exists a utility profile $u(s) \in \mathbf{U}(s)$, such that $e_i(s) = u_{-i}(s)$ for all player i; and for all $s \notin \text{supp}(\sigma^*)$ beliefs satisfy $e_i(s) = u_{-i}$ where $(u_{-i}, u_i) \in \mathbf{U}(s)$ for some u_i, for all player i. When beliefs are weakly consistent, players may disagree about the others' welfare outside of equilibrium, since their joint beliefs might not solve the interdependent utility system $\mathcal{U}$, although individual beliefs are justifiable. Adjusting Definition 1 so that beliefs are weakly consistent yields a *weak equilibrium* with an associated outcome set $\mathcal{O}^{**}$.

Proposition 2 clarifies to what extent the modified consistency condition affects the predictive power of the empathetic framework. The next result does not require Assumption 1.

Proposition 2 (Outcome Equivalence). *The equilibrium sets are equivalent $\mathcal{O}^* = \mathcal{O}^{**}$.*

Consistent beliefs are weakly consistent, and thus $\mathcal{O}^* \subset \mathcal{O}^{**}$. The other direction is more involved, as seen in Appendix A.1. For an intuition, consider a pure (weak) equilibrium outcome $(s^*, v^*) \in \mathcal{O}^{**}$. Notice that, for any potential deviation s_i of player i, player j's utility $U_j(s_i, s^*_{-i}, e_j(s_i, s^*_{-i}))$ and beliefs $e_j(s_i, s^*_{-i})$ are irrelevant for assessing whether s^*_j is optimal for player j, given s^*_{-j} — as player j only cares about unilateral deviations. Thus, for all strategy profiles $s \neq s^*$, we can construct consistent beliefs from weakly consistent beliefs, ensuring that s^* remains a mutual best response. The takeaway point is that weakening the consistency condition has no impact in terms of equilibrium outcomes, which may be surprising. This result not only extends our theory to less restrictive settings, but also provides a cornerstone of the characterization of equilibrium outcomes in section §5.

We close this section by making an observation regarding the computation of equilibria in empathetic games. Interestingly, one can find the equilibrium outcome set by following a simple decomposition of the utility correspondence $\mathbf{U}$. First, take a selection of this correspondence, namely, a single-valued function $u^r : \mathcal{S} \to \mathbb{R}^I$ satisfying $u^r(s) \in \mathbf{U}(s)$ for all profiles $s \in \mathcal{S}$. Next, find the set of Nash equilibria in the reduced-form game $\langle I, \mathcal{S}, u^r \rangle$. By Definition 1, the equilibria of this reduced-form game remain equilibria of the overall empathetic game. Finally, repeat this procedure for all distinct selections u^r. Then, one can, typically, construct the equilibrium set $\mathcal{O}^*$ in finitely many steps by Proposition 1 — for generically, there is a finite number of selections u^r. In §A.3, we show that *the set of equilibria is the union of equilibria in the reduced-form games* (Claim A.1.1).

Remark 1 (On Generic Finiteness of Equilibria). The smoothness requirement in Assumption 1 allows us to use differential topology and invoke transversality theorems (Milnor, 1997). For a standard normal form game, Wilson (1971) shows that, generically, there can be at most a finite number of equilibria. Wilson's inductive argument — an extension of the Lemke and Howson algorithm (Lemke and Howson, 1964) — requires that each player's payoff at any strategy profile can be independently perturbed. In empathetic games, perturbations are over the utility system, rather than the reduced-form utilities (i.e., elements of the utility set $\mathbf{U}$), and so Wilson's construction is not applicable. Our proof uses an alternative argument that relies on a transversality result. Likewise, we cannot use a version of Sard's theorem for generalized equations (Theorem 4.1 in Reinoza (1983) invoked, e.g., by Gül et al. (1993)) to argue generic regularity of mixed equilibria in a normal form game, because it assumes that the system of equations can be perturbed by exogenous parameters (here p). Here, transversality is shown with respect to perturbations of exogenous and endogenous variables (p and u), which introduces novel technical challenges. $\Diamond$

Remark 2 (On Efficiency of Equilibria). In empathetic games, we can perform normative analysis if we focus on outcomes instead of strategy profiles. An outcome (s, v) *Pareto dominates (is dominated by)* (s', v') if $v \geq v'$ $(v' \geq v)$ with strict inequality for at least one player. So an outcome is *Pareto efficient* if it is not dominated by any other outcome. In general, *the set of Pareto efficient outcomes is non-empty* (Claim A.1.1).

Inefficiencies may arise in this context not only because players choose socially-suboptimal strategies, but also because of the emotional reinforcement process that may lead agents to low welfare levels (see Example A.1.1). In the intergenerational altruism literature, reduced-form utilities over consumption streams are unique and may be time-inconsistent, generating alternative sources of inefficiencies (Saez-Marti and Weibull, 2005; Pearce, 2008). Recently, Ray and Vohra (2019) show that, if players are affected only by their own action and the realized payoffs of others, then all equilibria are pareto efficient, provided the interdependent utility system $\mathcal{U}$ yields unique and stable reduced-form payoffs. $\diamondsuit$

5 Characterization of Equilibrium Outcomes

We now provide a characterization of a set of equilibria. Assume that players are "cautious" when assessing hypothetical deviations, meaning that whenever their realized utilities can take multiple values for a given strategy profile, they expect their worst outcome to prevail.[17]

More precisely, define the justifiable utility set for player i, given strategy profile s, as $\mathbf{U}_i(s) \equiv \{u_i : (u_i, u_{-i}) \in \mathbf{U}(s) \text{ for some } u_{-i}\}$. Player i's *maxmin utility function* $\underline{v}_i : \sum_{-i} \to \mathbb{R}$ is the maxmin utility that player i can achieve when others play σ_{-i}:

$$\underline{v}_i(\sigma_{-i}) \equiv \sup_{s'_i \in S_i} \sum_{s_{-i} \in \mathcal{S}_{-i}} \inf \mathbf{U}_i(s'_i, s_{-i}) \sigma_{-i}(s_{-i}) \tag{3}$$

Observe that (3) is player i's best-response payoff to σ_{-i}, assuming the worst-case scenario for player i. These are the most favorable beliefs to support σ as an equilibrium profile. Indeed, letting $\underline{v}(\sigma) \equiv (\underline{v}_i(\sigma_{-i}))_{i \in I}$ we find:

Proposition 3. *A tuple (σ^*, v^*) is an equilibrium outcome if and only if* (C.1) *all players get at least their maxmin utility, i.e., $v^* \geq \underline{v}(\sigma^*)$; and* (C.2) *for all player i, any strategy s_i played with positive chance yields utility v^*_i when others play σ^*_{-i}, for some weakly consistent beliefs.*

First, observe that condition (C.1) holds iff for every player i and strategy s_i not in the support of σ^*_i, we have $v^*_i \geq \sum_{s_{-i}} \inf \mathbf{U}_i(s_i, s_{-i}) \sigma^*_{-i}(s_{-i})$. Second, Proposition 3 allows us to easily identify which profiles cannot be an equilibrium. By condition (C.1), any strategy profile that induces a utility less than the maxmin utility level for all players cannot be

[17] These beliefs are weakly consistent; thus, we can use them to characterize equilibria using Proposition 2.

implemented as an equilibrium. Thus, as a corollary, characterizing pure equilibrium outcomes is very simple, as condition (C.2) trivially holds. A tuple (s^*, v^*) is a pure equilibrium outcome if and only if all players get at least their maxmin utility, i.e., $v^* \geq \underline{v}(s^*)$, where slightly abusing notation, $\underline{v}(s)$ is the maxmin utility profile when s is played with probability 1. Finally, for "totally mixed" strategies, i.e., $\sup(\sigma^*) = \mathcal{S}$, condition (C.1) immediately obtains, and thus only condition (C.2) needs to be checked.[18]

In the next section, we provide examples that illustrate the usefulness of Proposition 3.

6 Games of Sympathy and Antipathy

Pre-existing relationships among individuals can affect how individuals empathize, i.e., how they perceive and experience the emotions of others both positively and negatively. Empathy can manifest as either sympathy or antipathy. Indeed, according to De Waal (2008): "In human studies, subjects tend to sympathize with a confederate's pleasure or distress when they perceive the relationship as cooperative, and yet show an antipathetic response (distress at seeing others' pleasure, or pleasure at seeing others' distress) if they perceive the relationship as competitive" (p. 291).

6.1 Sympathetic Games

Social psychology demonstrates that sympathy is often observed in cooperative settings, where the interests of the involved parties are aligned, such as in workplaces (Zillman and Cantor, 1977; Lanzetta and Englis, 1989; De Waal, 2008).[19] Also, from an evolutionary viewpoint, sympathetic attitudes naturally arise among subjects with similarity, familiarity, social closeness, and common experiences (Batson, 2011). These aspects trigger emotional contagion, where agents feel what others feel without expecting anything in return. Notice that this phenomenon is in contrast to reciprocity theory, where agents want to get even with other agents. In reciprocity theory, agents view other agents favorably and unfavorably, depending on the specific conditions present.

In this section, consider players whose realized utilities are positively related, i.e., they are sympathetic toward each other. A *sympathetic game* is an empathetic game in which for every player $i \in I$, the utility function $U_i(s, \cdot)$ is increasing for all profiles $s \in \mathcal{S}$ and strictly

[18] One could also provide an upper-bound of possible equilibrium payoffs. Indeed, define player i's maxmax utility function by changing the "inf" for a "sup" in equation (3), resembling optimistic beliefs. This would yield a vector of payoffs, say, $\bar{v}(\sigma) \equiv \bar{v}_i(\sigma_{-i})_{i \in I}$. Clearly, if (v, σ) is an equilibrium outcome, then $v \leq \bar{v}(\sigma)$.

[19] Preston and De Waal (2002) survey the literature on intensity of empathy in humans (animals).

increasing for some $s \in \mathcal{S}$.[20]

Exploiting the extra structure of sympathetic games, we show in §A.3 that *the set* $\mathbf{U}(s)$ *is a complete lattice for all profiles* s, and thus $\mathbf{U}(s)$ has a minimal and maximal element. Moreover, *in games with two players, the utility set* $\mathbf{U}(s)$ *is totally ordered* (Claim A.3.1). These results formally capture the idea that, due to emotional contagion among players, similar behavior can lead to either joint excitement or shared frustration. This insight appears particularly relevant for managerial practices, because it illustrates a psychological force that may be important in analyzing how to keep workers' morale high (Bewley, 1999).[21]

In Example 3.2-A, notice that $\underline{v}_i(\sigma_j) = 1$ for all i, j, and so (C, G) and (G, C) cannot be equilibrium profiles, by Proposition 3. However, (C, C) and (G, G) can be supported as equilibrium outcomes, provided beliefs $e_i^*(C, C) = 2$ for all player i. In particular, the totally mixed strategy $\sigma_i(G) = 2/3$ is an equilibrium, since strategies C and G yield the same utility for all i, j (condition C.2).[22] This example suggests that, although agents derive pleasure from others' pleasure, and thus collaborating (or going to the date) is efficient, an "environment" that triggers positive self-reinforcing beliefs among the agents also appears to be useful to ensure an efficient outcome.

We next discuss the marginal effects of sympathy. One can imagine that people are more sympathetic toward those who are less fortunate. For instance, a person who is sick may elicit more sympathy (per utile) than a person who is healthy. This demands that the marginal increase in sympathy lessens as the welfare of others rises. Also, sympathetic individuals may exhibit extreme compassion at the margin for those who are in peril, but have no compassion for the auspicious ones. We say that a utility system $\mathcal{U} = (U_i)_{i \in I}$ exhibits *diminishing (rising) marginal sympathy* at s if $U_i(s, \cdot)$ is strictly concave (convex) for any player i; and it satisfies the *Inada* conditions if system $\mathcal{U}$ is continuously differentiable and $\lim_{u_j \downarrow -\infty} \partial U_i(s, u_j)/\partial u_j = -\infty$ and $\lim_{u_j \uparrow \infty} \partial U_i(s, u_j)/\partial u_j = 0$, for every player i. Since any strictly concave function $U_i : \mathbb{R} \to \mathbb{R}$ is necessarily unbounded, we now relax Assumption 1.

[20]Consider the component-wise order, so that for any pair of vectors $u, u' \in \mathbb{R}^I$, $u \geq u'$ iff $u_i \geq u'_i$ for all $i \in I$ and $u > u'$ if for some i inequality is strict. We say that $U_i(s, \cdot)$ is *increasing at* s if $u_{-i} \geq u'_{-i}$ implies $U_i(s, u_{-i}) \geq U_i(s, u'_{-i})$ and *strictly increasing* if for $u_{-i} > u'_{-i}$ one has $U_i(s, u_{-i}) > U_i(s, u'_{-i})$. Analogous definitions hold for decreasing and strictly decreasing functions.

[21]The structure of the utility set $\mathbf{U}(s)$ for sympathetic games share formal similarities with the set of Nash equilibria in games with strategic complementarities — namely, both sets contain a largest and smallest element — as their fixed-point correspondences are increasing. However, from a behavioral perspective, this similarity breaks down: a sympathetic game may give rise to a game with strategic substitutes.

[22]Notice that, in sympathetic games, when utility functions are symmetric at a given strategy profile s, the utility set $\mathbf{U}(s)$ takes a particularly simple form, namely, a utility profile $u \in \mathbf{U}(s)$ iff $u_i = u_j$ for all players i, j. This means that the utility set $\mathbf{U}(s)$ can be determined as the solutions of a single equation $u_i = U_i(s, u_i, ..., u_i)$. To see this, suppose that an element of $\mathbf{U}(s)$ is not symmetric. Then, there must exist a player i such that $u_i \leq u_{i'}$ for all i' with strict inequality for some j. Since u_{-i} and u_{-j} differ only in one element that is higher in u_{-i} than u_{-j}, we have $u_i = U_i(s, u_{-i}) \geq U_i(s, u_{-j}) = u_j$, which is a contradiction.

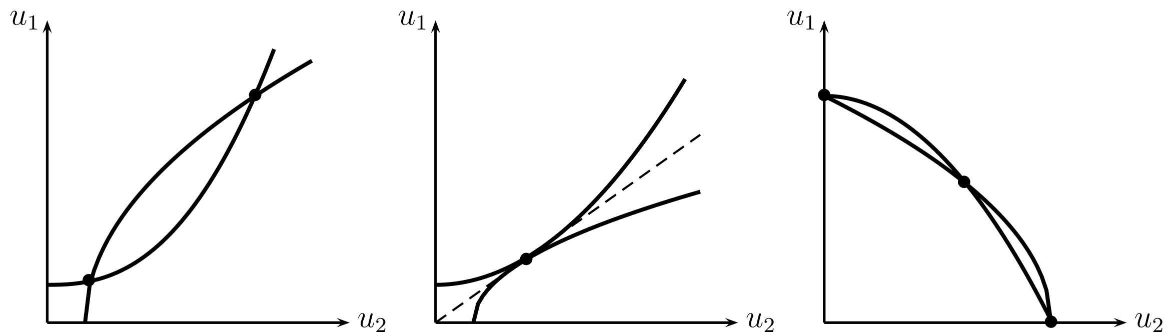

Figure 1: **Understanding the utility set U.** LEFT: When utility functions exhibit diminishing sympathy, the utility function for player 2 is increasing and concave, and thus its inverse is increasing, but convex. Thus, in the (u_1, u_2)-space, the utility of player 1 is increasing and concave, while the inverse payoff for player 2 is increasing and convex. These functions intersect twice. MIDDLE: A unique fixed point is non-generic, provided Inada. RIGHT: With rising marginal antipathy, player 1's utility function is decreasing and concave and so is the inverse of player 2's utility function; thus, these functions can intersect more than twice.

Diminishing sympathy and Inada are natural behavioral assumptions for preferences that are consistent with usual economic logic. Yet a unique solution to a payoff system, often assumed in the literature, is non-generic.[23] As seen in Example 3.2 and Figure 1, *in two-player sympathetic games, the utility set* $\mathbf{U}(s)$ *has, at most, two elements, provided that players exhibit monotonic marginal sympathy.* This holds generically when the utility system satisfies the Inada conditions (Claims A.3.2–A.3.3).

Finally, in the left panel of Figure 1, the two solutions to the utility system are not equally plausible in light of the tatonnement process (2) described in §3. Clearly, for all initial beliefs above the highest solution, the belief process converges to this solution which is, thus, stable. Yet, for starting beliefs below the lowest solution, feedback effects generates explosive dynamics that diverges to $-\infty$ for both players (extreme misery). Only for the knife-edge case in which initial beliefs coincide with the lowest solution, that solution can persist over time. In this sense, the lowest solution is less likely to prevail.

6.2 Antipathetic Games

Antipathy is the opposite of sympathy — a feeling of dislike for someone. Lanzetta and Englis (1989) and Zillman and Cantor (1977) show that antipathy often arises in competitive environments, where one party's gains result in losses for the other party. As Bertrand Russell once wrote, "I care for very few people and have several enemies—two or three at least whose pain is delightful to me" (Russell, 2000, p. 73). In these environments, the emotional contagion process induces realized utilities to be negatively related. Following §6.1, an *an-*

[23]The Inada assumption is critical. Suppose that $U_i(s, u_i) = u_i - \exp(-u_j)$ for $i = 1, 2$, so U_i is strictly increasing concave, and $\lim_{u_j \uparrow \infty} \partial U_i(s, u_j)/\partial u_j = 1$. For small perturbations, there is a unique fixed point.

tipathetic game is an empathetic game in which for every player $i \in I$, the utility function $U_i(s, \cdot)$ is decreasing for all profiles $s \in \mathcal{S}$ and strictly decreasing for some $s \in \mathcal{S}$

Unlike in sympathetic games, the utility set $\mathbf{U}(s)$ does not have an ordered structure with more than two players unless strong parametric assumptions for $\mathcal{U}$ are present, such as symmetric linearity. This highlights that, in antipathetic games, emotions are not transitive, which adds another layer of complexity. More precisely, consider an antipathetic game with three players $i = 1, 2, 3$. An exogenous increase in the payoff of player 1 has a direct adverse effect on player 3's payoff. It also reduces player 2's payoff, which indirectly improves player 1's payoff. If antipathy between players $1, 2$ and $2, 3$ is strong, while the same between players 1 and 3 is only mild, then the indirect effect on player 3's payoff may dominate. In other words, while player 1 dislike player 3, the increase of player 3's payoff has a strictly positive impact on player 1. In this sense, the ancient proverb applies: "The enemy of my enemy is my friend." This logic does not arise in sympathetic games, as positive emotions are transitive, or "The friend of my friend is my friend." In antipathetic games, Henceforth, we focus on the direct effects of antipathy, restricting our attention to two-player antipathetic games.

In general, in two-player antipathetic games, for any strategy profile, there exists a utility profile that is the best for one player and the worst for the other. In fact, *this is a general property of two-player antipathetic games* (Claim A.3.4).[24] Also, unlike sympathy, antipathetic games with concave utility functions (rising antipathy) do not limit the number of elements of the utility system — as seen in the right panel of Figure 1.

Suppose Alice and Bob dislike each other, so that if they both choose G, then each player gets $U_i(s, u_j) = \sqrt{4 - 2u_j}$ if $u_j \leq 2$ and $U_i(s, u_j) = 0$ otherwise. Payoffs to other profiles are as in Example 3.2. Observe that $\mathbf{U}(G, G) = \{(0, 2), (1.23, 1.23), (2, 0)\}$.[25] So aside from the asymmetric solutions, in which one player enjoys the misfortune of the other, there is an instance in which both Anne and Bob receive the same payoff.[26] Here, the maxmin utilities for pure strategies obey $\underline{v}_i(G) = \underline{v}_i(C) = 1$. Thus, (G, C) and (C, G) cannot be supported as equilibria, by Proposition 3. Yet, (G, G) can be supported as equilibrium *only* for beliefs that coordinate on $u_i(G, G) = 1.23$. Profile (C, C) can always be supported as an equilibrium, using the reasons previously discussed. Finally, notice that $\underline{v}_i(\sigma_j) = 1$ for any σ_j, and so the only equilibrium in mixed strategy is symmetric and entails $\sigma_i(G) = 0.9$ and beliefs that coordinate on $u_i(G, G) = 1.23$. Altogether, this example suggests that the only stable way

[24] In two-player antipathetic games, every utility profile in $\mathbf{U}(s)$ is "payoff efficient" in the sense that there is no other utility profile in $\mathbf{U}(s)$ that Pareto dominates (Remark 2) the other in terms of welfare. As a result, inefficiencies stem exclusively from the suboptimal choice of players' strategies. By contrast, in sympathetic games this is usually not the case, as utility profiles are positively related and ordered (Claim A.3.1).

[25] Unlike sympathy §6.1, here the elements of $\mathbf{U}(s)$ are not symmetric even if utility functions $(U_i)_i$ are.

[26] This is a general feature of empathetic games with symmetric utility functions (Observation A.3.1).

to put together two people that dislike each other is by neutralizing their emotions such that no one can benefit from the dissatisfaction of the other, provided their outside options are neither too high nor too low.

7 Paternalistic and Intention-based Preferences

We now study the relationship of our framework to other forms of interdependent preferences.

A. Outcome-based/Paternalistic Preferences. Our paper relates to the literature on *material* games, which are commonly used to model interdependent preferences; see Sobel (2005) for a survey. This approach exploits the standard game-theoretic formulation, wherein utilities are a function of outcomes. In material games, an outcome is a distribution of material payments, such as consumption or money, across players. Specifically, for a fixed game, an allocation rule assigns material payoffs $x(s) \equiv (x_i(s))_i$ when profile s is played. Preferences are represented by a compound utility function $V_i(x(\cdot))$. Player i is deemed as *paternalistically altruistic* towards player j if his utility rises in the material payoff of j (i.e., x_j). This approach is flexible as preferences can be easily tweaked to rationalize experimental data that is otherwise inconsistent using models that assume that agents are purely selfish ($V_i(x(\cdot)) \equiv V_i(x_i(\cdot))$). Versions of this specification have been used previously, e.g., in Levine (1998); Fehr and Schmidt (1999); Bolton and Ockenfels (2000) and Grohn et al. (2014).

However, because of the reduced-form nature of this approach, it is unclear why and how these utilities depend on the distribution of material payoffs, or what determines the specific shape of these utility functions. If we interpret $V_i(x(\cdot))$ as primitive utility functions, then we may encounter paradoxes. As we show in §3.2-B, in certain settings a player that is paternalistically altruistic towards another player may take actions that could indeed hurt this player, as players do not internalize how their actions impact the final welfare of others.

Alternatively, we could interpret $V_i(x(\cdot))$ as a realized utility function coming from a primitive interdependent utility system $\mathcal{U}$. Suppose players care about both their own material outcome and the welfare of others so that $U_i\,(x_i, u_j) = x_i + b\sum_{j\neq i} u_j$. This utility system admits a paternalistic representation $(V_i(x))_i$ when $b \neq 1/(I-1)$ and $b \neq -1$, which is given by:

$$V_i(x) = \frac{1+2b-bI}{(1+b)(1+b-bI)} x_i + \frac{b}{(1+b)(1+b-bI)} \sum_{j\neq i} x_j. \tag{4}$$

Notice that, since players are affected by both their own material gain and the total welfare of others, we can think of players as if they cared about a linear combination of their own material gains and those of others.[27] Also, observe that the coefficients of V_i summarize those

[27]Bramoullé (2001) and Bourlès et al. (2017) study how primitive relationships of agents' utilities impact

of the primitive utility system U_i in an interesting way. For sympathetic games (i.e. $b>0$), the weight placed on others' material payoffs in (4) is positive iff sympathy among the players is not too strong, or $b < 1/(I-1)$.[28] That is, reduced-form payoffs may fall in others' material payoffs when agents are too sympathetic, $b > 1/(I-1)$.[29,30] In general, the non-paternalistic framework clarifies when (reduced-form) preferences that increase in the others' material outcomes are indeed capturing a genuine concern for the others' final welfare.

B. INTENTION-BASED PREFERENCES. Finally, our paper relates to the psychological games literature (Geanakoplos et al., 1989; Dufwenberg, 2008; Rabin, 1993), wherein players want to be kind and unkind to whoever is kind and unkind to them, respectively. In psychological games, utility functions not only depend on outcomes but also on higher-order beliefs, which are beliefs about beliefs about beliefs about choices. For example, in Rabin (1993), player's i utility function is $V_i(x(s)|\hat{s}) \equiv v_i(x_i(s)) + \alpha_i(\hat{s})v_j(x_j(s))$, where v_i, v_j are material utilities that depend on the allocation rule $x(s)$, $\alpha(\hat{s})$ measures how much player i cares for j, and $\hat{s} = (\hat{s}_i, s_j)$ denotes player i beliefs about what player j believes about him.[31] The equilibrium concept discussed in Geanakoplos et al. states that, given high order beliefs, players must play a best response and, given those responses, high order beliefs must be justified by their play. In the previous example, this means that if s is an equilibrium profile, then it must be a mutual best response, given $\hat{s}$, and $s = \hat{s}$. By putting more structure on $\alpha(\cdot)$ one can model situations in which players, say, reciprocate kindness with kindness and meanness with meanness; see Charness and Rabin (2002) and Falk and Fischbacher (2006).[32]

In intention-based models, agents have a strategic reason to reciprocate behavior. This assumption may be better suited for environments in which players are anonymous, and thus affective empathy is less important. However, in other settings, players may know each other in such a way that their welfare are genuinely interlinked (e.g., family and relatives, friends, or enemies), or interactions may be face-to-face. In these settings, affective empathy and emotional contagion are, indeed, relevant, as the psychological evidence shows; see §1.

Finally, at a more technical level, in our paper utility functions depend on both strategies

reduced-form relationships in linear and separable environments; see Theorem 2 in Bourlès et al. (2017).

[28] For altruistic interdependent utility systems, Pearce (2008), Bergstrom (1999), and Bramoullé (2001) provide sufficient conditions so that paternalistic payoffs $V_i(x)$ are well defined and increasing in others' material payoffs. In particular, conditions (A1)–(A3) in Pearce (2008) are equivalent here to $b \in (0, 1/(I-1))$.

[29] This situation is unstable as feedback effects lead to explosive utilities that diverge to either ∞ or $-\infty$.

[30] Bernheim (1989) and Bernheim and Stark (1988) discuss how sympathy can lead to paradoxical results.

[31] Segal and Sobel (2007) provide conditions under which a player's preferences over strategies s can be represented as a weighted average of the utility from outcomes of the individual and his opponents. The weight one player places on an opponent's utility from outcomes depends on the players' joint behavior $\hat{s}$.

[32] Levine (1998) and Gul and Pesendorfer (2016) provide an alternative (epistemic) approach to model intentions using behavioral types which encode observable characteristics and personality traits. Winter et al. (2016) develops a theory of strategic emotions in which players simultaneously choose strategies and mental states to best respond one another, while mental states determine preferences over outcomes.

and the realized utilities of others. Because players act independently, reduced-form utilities over outcomes are endogenously-determined through beliefs about the others' realized utilities. Like Geanakoplos et al. (1989), we also follow an equilibrium approach and use a consistency condition. However, in our setting players internalize how their deviations would impact the utility of others and thereby their own utility. Thus, our notion of consistency applies to all strategy profiles and not just equilibrium ones.[33]

8 Concluding Remarks

Empathy shapes many, if not most, social interactions. In this paper, we propose a framework that captures a type of empathy that has been extensively documented in social psychology but unexplored in the economics literature. We focus on the role of affective empathy and the related emotional contagion process among players. In our framework, players care not only about a chosen strategy profile, but also about others' realized utilities; thus, our theory crucially distinguishes between primitive-utility and realized-utility functions. To capture emotional feedback effects, we allow realized utilities to be interdependent. This assumption raises conceptual and technical obstacles. Because feedback effects may lead to multiple realized utilities, one can think of these games as if players' preferences are described by correspondences instead of utility functions. We provide a parsimonious and tractable solution concept and characterize the corresponding set of equilibrium outcomes. We also provide examples that illustrate the scope of the theory.

Our framework is not only tractable, but also useful in explaining psychological and behavioral phenomena. We are currently exploring the role of affective empathy in principal-agent settings. In particular, in Vásquez and Weretka (2016), we consider a labor market in which a manager chooses both a team of workers and their compensation. The main innovation is to introduce affective empathy among workers in the workplace, giving rise to novel sources of multiplicity. We explore how firms respond to productivity shocks and show that the model rationalizes the empirical findings of Bewley (1999), which establish that in recessions firms are reluctant to reduce their employees' wages, as wage reductions can damage the workers' *morale*. This empirical fact has been hard to reconcile with alternative formulations of pro-social preferences. This indicates that our paper may provide a natural stepping stone towards a more general understanding of the role of *morale* in economics.

In this paper, we consider the simplest case of simultaneous move games. This allows us to

[33] It is immediate to adapt our solution concept to settings with "limited empathy," in which players expect an equilibrium utility profile u^* and best respond to one another taking u^* as given. In this case, beliefs need to be consistent only for equilibrium profiles, like in Geanakoplos et al. (1989).

study strategic interactions with pre-existing relationships (e.g., sympathy or antipathy). An important direction for future research is to allow agents' relationships to evolve depending on *how* the game transpires. For this goal, a natural first step is to extend the framework to encompass dynamic considerations, and then focus on how pro-active behaviors shape mutual attitudes. Also, allowing dynamics seems important for experimental work, because most of the designs in this field have a sequential-move protocol.

Finally, in our current setting, we isolate a psychological force and let emotions transmit unconsciously or subconsciously among players. It would be interesting to merge our approach with psychological games to also capture strategic emotion transmission. For example, players can be either sympathetic or antipathetic, depending on how certain behavior is perceived.[34] Allowing utility functions to depend not only on others' realized utilities, but also on hierarchies of beliefs regarding behavior, would provide a natural language to study strategic and unconscious emotion contagion and its behavioral implications. Other conceptual issues that remain open include exploring other contending solution concepts (e.g., those based on learning or rationality), and axiomatic foundations for interdependent utilities.[35]

References

Andreoni, J. (1989): "Giving with impure altruism: applications to charity and Ricardian equivalence," *The Journal of Political Economy*, 1447–1458.

Barro, R. J. (1974): "Are government bonds net wealth?" *Journal of Political Economy*, 82, 1095–1117.

Batson, C. D. (2011): *Altruism in humans*, Oxford University Press, USA.

Becker, G. S. (1974): "A theory of social interactions," *Journal of political economy*, 82, 1063–1093.

Bergstrom, T. (1989): "Puzzles: Love and spaghetti, the opportunity cost of virtue," *Journal of Economic Perspectives*, 3, 165–173.

Bergstrom, T. C. (1997): "A survey of theories of the family," *Handbook of population and family economics*, 1, 21–79.

——— (1999): "Systems of benevolent utility functions," *Journal of Public Economic Theory*, 1, 71–100.

[34] Indeed, neuroscientific experiments show that seeing the pain of a cooperative party activates brain areas related to pain, while seeing the pain of a competitor stimulates reward-related areas (Singer et al., 2006).

[35] Galperti and Strulovici (2017) develop an axiomatic theory of altruism in an intergenerational context.

Bernheim, B. D. (1989): "Intergenerational altruism, dynastic equilibria and social welfare," *The Review of Economic Studies*, 56, 119–128.

Bernheim, B. D. and O. Stark (1988): "Altruism within the family reconsidered: Do nice guys finish last?" *The American Economic Review*, 1034–1045.

Bewley, T. F. (1999): *Why wages don't fall during a recession*, Harvard University Press.

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

Bourlès, R., Y. Bramoullé, and E. Perez-Richet (2017): "Altruism in networks," *Econometrica*, 85, 675–689.

Bramoullé, Y. (2001): "Interdependent utilities, preference indeterminacy, and social networks," *mimeo*.

Charness, G. and M. Rabin (2002): "Understanding social preferences with simple tests," *The Quarterly Journal of Economics*, 817–869.

Courty, P. and M. Engineer (2017): "A pure hedonic theory of utility and status: Unhappy but efficient invidious comparisons," *mimeo*.

Dasgupta, P. (2008): "Discounting climate change," *Journal of Risk and Uncertainty*, 37, 141–169.

De Waal, F. B. (2008): "Putting the altruism back into altruism: The evolution of empathy," *Annual Review of Psychology*, 59, 279–300.

Dufwenberg, M. (2008): "Psychological games," *The New Palgrave Dictionary of Economics*, 2.

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

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

Fels, S. and R. Zeckhauser (2008): "Perfect and total altruism across the generations," *Journal of Risk and Uncertainty*, 37, 187–197.

Fudenberg, D. and D. K. Levine (1993): "Self-confirming equilibrium," *Econometrica*, 523–545.

Galperti, S. and B. Strulovici (2017): "A theory of intergenerational altruism," *Econometrica*, 85, 1175–1218.

Geanakoplos, J., D. Pearce, and E. Stacchetti (1989): "Psychological games and sequential rationality," *Games and Economic Behavior*, 1, 60–79.

Grohn, J., S. Huck, and J. M. Valasek (2014): "A note on empathy in games," *Journal of Economic Behavior and Organization*, 108, 383–388.

Gül, F., D. Pearce, and E. Stacchetti (1993): "A bound on the proportion of pure strategy equilibria in generic games," *Mathematics of Operations Research*, 18, 548–552.

Gul, F. and W. Pesendorfer (2016): "Interdependent preference models as a theory of intentions," *Journal of Economic Theory*, 179–208.

Hatfield, E., L. Bensman, P. D. Thornton, and R. L. Rapson (2014): "New perspectives on emotional contagion: A review of classic and recent research on facial mimicry and contagion," *Interpersona*, 8, 159.

Hatfield, E., J. T. Cacioppo, and R. L. Rapson (1993): "Emotional contagion," *Current Directions in Psychological Science*, 2, 96–100.

Hawk, S. T., A. H. Fischer, and G. A. Van Kleef (2012): "Face the noise: Embodied responses to nonverbal vocalizations of discrete emotions." *Journal of Personality and Social Psychology*, 102, 796.

Iacoboni, M. (2009): "Imitation, empathy, and mirror neurons," *Annual Review of Psychology*, 60, 653–670.

Jackson, P. L., A. N. Meltzoff, and J. Decety (2005): "How do we perceive the pain of others? A window into the neural processes involved in empathy," *Neuroimage*, 24, 771–779.

Kalbe, E., M. Schlegel, A. T. Sack, D. A. Nowak, M. Dafotakis, C. Bangard, M. Brand, S. Shamay-Tsoory, O. A. Onur, and J. Kessler (2010): "Dissociating cognitive from affective theory of mind: a TMS study," *Cortex*, 46, 769–780.

Koopmans, T. C. (1960): "Stationary ordinal utility and impatience," *Econometrica*, 287–309.

LANZETTA, J. T. AND B. G. ENGLIS (1989): "Expectations of cooperation and competition and their effects on observers' vicarious emotional responses." *Journal of Personality and Social Psychology*, 56, 543.

LEMKE, C. E. AND J. T. HOWSON, JR (1964): "Equilibrium points of bimatrix games," *Journal of the Society for Industrial and Applied Mathematics*, 12, 413–423.

LEVINE, D. K. (1998): "Modeling altruism and spitefulness in experiments," *Review of Economic Dynamics*, 1, 593–622.

LINDBECK, A. AND J. W. WEIBULL (1988): "Altruism and time consistency: the economics of fait accompli," *Journal of Political Economy*, 96, 1165–1182.

MILNOR, J. W. (1997): *Topology from the differentiable viewpoint*, Princeton University Press.

NASH, J. (1950): "Equilibrium points in n-person games," *Proceedings of the National Academy of Sciences*, 36, 48–49.

PEARCE, D. G. (2008): "Nonpaternalistic sympathy and the inefficiency of consistent intertemporal plans," *Foundations in Microeconomic Theory*, 213.

PHELPS, E. S. AND R. A. POLLAK (1968): "On second-best national saving and game-equilibrium growth," *The Review of Economic Studies*, 35, 185–199.

PRESTON, S. D. AND F. B. DE WAAL (2002): "Empathy: Its ultimate and proximate bases," *Behavioral and Brain Sciences*, 25, 1–20.

RABIN, M. (1993): "Incorporating fairness into game theory and economics," *The American Economic Review*, 1281–1302.

RAMSEY, F. P. (1928): "A mathematical theory of saving," *The Economic Journal*, 38, 543–559.

RAY, D. (1987): "Nonpaternalistic intergenerational altruism," *Journal of Economic Theory*, 41, 112–132.

RAY, D. AND R. VOHRA (2019): "Games of love and hate," *Journal of Political Economy*, forthcoming.

REINOZA, A. (1983): "Global behavior of generalized equations: A Sard theorem," *SIAM Journal on Control and Optimization*, 21, 443–450.

Rockafellar, R. T. (1970): *Convex analysis*, Princeton University Press.

Russell, B. (2000): "Self-appreciation," *The collected papers of Bertrand Russell. Volume one: Cambridge essays*, 1, 73.

Saez-Marti, M. and J. W. Weibull (2005): "Discounting and altruism to future decision-makers," *Journal of Economic Theory*, 122, 254–266.

Segal, U. and J. Sobel (2007): "Tit for tat: Foundations of preferences for reciprocity in strategic settings," *Journal of Economic Theory*, 136, 197–216.

Singer, T. and E. Fehr (2005): "The neuroeconomics of mind reading and empathy," *The American Economic Review*, 95, 340–345.

Singer, T., B. Seymour, J. O'doherty, H. Kaube, R. J. Dolan, and C. D. Frith (2004): "Empathy for pain involves the affective but not sensory components of pain," *Science*, 303, 1157–1162.

Singer, T., B. Seymour, J. P. O'Doherty, K. E. Stephan, R. J. Dolan, and C. D. Frith (2006): "Empathic neural responses are modulated by the perceived fairness of others," *Nature*, 439, 466–469.

Smith, A. (1822): *The theory of moral sentiments*, vol. 1, J. Richardson.

Sobel, J. (2005): "Interdependent preferences and reciprocity," *Journal of Economic Literature*, 392–436.

Vásquez, J. and M. Weretka (2016): "Empathy, unemployment, and wages," *mimeo*.

Wilson, R. (1971): "Computing equilibria of n-person games," *SIAM Journal on Applied Mathematics*, 21, 80–87.

Winter, E., L. Méndez-Naya, and I. García-Jurado (2016): "Mental equilibrium and strategic emotions," *Management Science*, 63, 1302–1317.

Zillman, D. and J. R. Cantor (1977): "Affective responses to the emotions of a protagonist," *Journal of Experimental Social Psychology*, 13, 155–165.

A Omitted Proofs

A.1 Proofs of Section §4

Proof of Proposition 1 : First we demonstrate existence of equilibrium. Fix a profile $s \in \mathcal{S}$. Since the interdependent utility system $\mathcal{U}(s,\cdot)$ is bounded, there exists a closed box[36] $B \subset \mathbb{R}^I$, such that $\mathcal{U}(s,\mathbb{R}^I) \subset B$. Notice that $\mathcal{U}(s,\cdot)$ restricted to B maps into itself; the set B is convex and compact, and the map is continuous, by Assumption 1. Thus, by Brouwer Fixed Point Theorem, there exists $u \in B$ with $\mathcal{U}(s,u) = u$. This holds for any $s \in \mathcal{S}$, and so the utility profile set is non-empty, $\mathbf{U}(s) \neq \emptyset$, for any $s \in \mathcal{S}$.

Fix an empathetic game $\Gamma = \langle I, \mathcal{S}, \mathcal{U} \rangle$ with $\mathcal{U}$ satisfying Assumption 1. Let $u^r : \mathcal{S} \to \mathbb{R}^I$ be a *selection* of $\mathcal{U}(\cdot)$, namely, a function satisfying $u^r(s) \in \mathbf{U}(s)$ for all s. By non-emptiness of $\mathbf{U}(\cdot)$, at least one such selection exists. Next, consider a standard normal form game $\Gamma^r \equiv \langle I, \mathcal{S}, u^r \rangle$. By Nash (1950), the game Γ^r has a mixed strategy equilibrium σ^*. Next, define $e_i^*(s) = u_{-i}^r(s)$ for all s and i. Since $u^r(s) \in \mathbf{U}(s)$, we have that $u_i^r(s) = U_i(s, u_{-i}^r(s)) = U_i(s, e_i^*(s))$, and so for any (i, σ_i), i's reduced form payoff obeys:

$$U_i^R(\sigma|e_i^*) = \sum_{s\in\mathcal{S}} \prod_{j\in I} \sigma_j^*(s_j) u_i^r(s) \geq \sum_{s\in\mathcal{S}} \sigma_i(s_i) \prod_{j\neq i} \sigma_j^*(s_j) u_i^r(s) = U_i^R(\sigma_i, \sigma_{-i}^*|e_i^*)$$

where the inequality holds, for σ^* is a Nash equilibrium in Γ^r. Finally, Definition 1-i) holds, for e^* is jointly consistent for all $s \in \mathcal{S}$ by construction. Definition 1-ii) holds as σ^* is a Nash equilibrium in Γ^r. Thus, (σ^*, e^*) is an empathetic equilibrium of Γ. Using equation (1), and letting $v_i^* \equiv U_i^R(\sigma^*|e_i^*)$ for all i, we see that $(\sigma^*, v^*) \in \mathcal{O}^*$.

Next we demonstrate generic finiteness of equilibria. We develop a series of lemmas, which we then use to prove the main result.

Lemma 1. *For any open set $\mathcal{P}$, there exists a subset of perturbations $\mathcal{P}^0$ with full Lebesgue measure such that, in any game Γ_p with $p \in \mathcal{P}^0$, the utility set $\mathbf{U}(s)$ is finite for all $s \in \mathcal{S}$.*

Proof: Consider a function $g : \mathbb{R}^{I\times|\mathcal{S}|} \times \mathcal{P} \to \mathbb{R}^{I\times|\mathcal{S}|}$, where given $u = (u_s)_{s\in\mathcal{S}}$, $u_s \in \mathbb{R}^I$; and $p = (p_s)_{s\in\mathcal{S}}$, $p_s \in \mathbb{R}^I$, one has

$$g(u,p) \equiv (\mathcal{U}(s,u_s) - u_s + p_s)_{s\in\mathcal{S}} \tag{5}$$

The collection of all roots of function $g(\cdot,p)$ uniquely defines $\mathcal{U}(\cdot)$ in a perturbed empathetic game. We will show that $g(\cdot,p)$ has a finite number of roots for almost every perturbation

[36] A *closed box* $B \subset \mathbb{R}^I$ is a set $B \equiv \{b \in \mathbb{R}^I : \bar{b} \geq b \geq \underline{b}\}$ for some $\bar{b}, \underline{b} \in \mathbb{R}^I$ with $\bar{b}_i > \underline{b}_i$ for all $i = 1,\ldots,I$. Unlike a closed ball, a closed box has an orderedvstructure that we later exploit in §6.

p. First, since g is additive separable in p with each parameter $p_{i,s}$ perturbing one equation and $\mathcal{U}$ is continuously differentiable, function $g(\cdot)$ is smooth and so transverse to zero (i.e., $g \pitchfork 0$). Second, by the Transversality Theorem, there exists a set $\mathcal{P}^0 \subset \mathcal{P}$ with full Lebesque measure such that $g(\cdot, p) \pitchfork 0$ for all $p \in \mathcal{P}^0$. Thus, for each root $u \in g^{-1}(0, p)$, the Jacobian of g has full rank. Third, by the Inverse Function Theorem, there exists a neighborhood around each root u, such that $g(\cdot, p)$ (restricted to this neighborhood) is a bijection. Thus, there can be at most one solution to $g(\cdot, p) = 0$; i.e., root $u \in g^{-1}(0, p)$ is an isolated point. Finally, by Assumption 1, for any $p \in \mathcal{P}^0$, there exists a closed box $B_p \subset \mathbb{R}^{I \times |\mathcal{S}|}$ such that $g^{-1}(0, p) \subset B_p$. Since B_p is compact and $g^{-1}(0, p)$ is a collection of isolated points, $g^{-1}(0, p)$ is necessarily finite.[37] So for almost every perturbation, utility set $\mathbf{U}(\cdot)$ is finite. $\square$

Now we exploit the geometry of our problem. Observe that the set of mixed strategy profiles $\Sigma = \times_{i \in I} \Delta(S_i)$ is a *polyhedral* set. Thus, by Theorem 19.1 in Rockafellar (1970), the set of strategy profiles Σ has finitely many *faces*, that we index by $k = 1, \ldots, K$ and denote by F^k. Next, call $\tilde{F}^k$ to the *relative interior* of F^k (i.e., $\tilde{F}^k \equiv \text{ri}F$), and let L^k be the *affine hull* of F^k (i.e., $L^k \equiv \text{aff}F$); namely, the smallest affine subspace containing F^k. Finally, denote by $(L^k)^\perp$ the *orthogonal complement* of L^k.

By Theorem 18.2 in Rockafellar (1970), the profile set Σ is partitioned by the relative interior of its faces; that is, $\Sigma = \bigcup_{k=1}^K \tilde{F}^k$ where $\tilde{F}^k \cap \tilde{F}^{k'} = \emptyset$ for all $k \neq k'$. Thus, for any profile $\sigma \in \Sigma$, there exists a unique face F^k such that $\sigma \in \tilde{F}^k$.[38] Also, for any $\tilde{F}^k$, there exist subsets $(S_i^k)_{i \in I}$ with $S_i^k \subset S_i$ such that $\text{supp}(\sigma) = \times_{i \in I} S_i^k \equiv \mathcal{S}^k$ for all $\sigma \in \tilde{F}^k$.

Next, fix $k \in \{1, \ldots, K\}$. For any player $i \in I$, define a function $v_i(s_i|\cdot) : \mathbb{R}^{|\mathcal{S}^k|} \times L^k \to \mathbb{R}$, that for each $s_i \in S_i^k$, vector $u_i \in \mathbb{R}^{|\mathcal{S}^k|}$ and σ_{-i} assigns a real number,

$$v_i(s_i|u_i, \sigma) = \sum_{s_{-i} \in \mathcal{S}_{-i}^k} \prod_{j \neq i} \sigma_j(s_j) u_i(s_i, s_{-i}),$$

and $v_i(s_i|\cdot) = 0$, if $s_i \notin S_i^k$. For $\sigma \in \tilde{F}^k$ this function gives the expected utility of a pure strategy $s_i \in S_i^k$ to player i, given others' playing σ_{-i} and support payoff vector $u_i \in \mathbb{R}^{|\mathcal{S}^k|}$. For any $(u, \sigma) \in \mathbb{R}^{I \times |\mathcal{S}^k|} \times L^k$, we define $v_i(u, \sigma) \equiv (v_i(s_i|u, \sigma))_{s_i \in S_i}$ and $v(u, \sigma) \equiv (v_i(u, \sigma))_{i \in I}$.

Given a tuple (e, σ) with $\sigma \in \tilde{F}^k$, we say that a utility vector $u(s) \in \mathbb{R}^{I \times |\mathcal{S}^k|}$ *generates beliefs on the support* if it solves $u(s) \in \mathbf{U}(s)$ for all $s \in \mathcal{S}^k$ and $e_i(s) = u_{-i}(s)$ for each $i \in I$.

[37]For if not, one then could find an infinite sequence of distinct solutions belonging to a compact set. This sequence would have a convergent subsequence, and so by continuity, the limit would be a root of $g(\cdot)$. But then, for any open neighborhood about this limit, we could find another root in this neighborhood, which contradicts the fact that roots are isolated.

[38]For example, a unit simplex Σ in $\mathbb{R}^3$ has *seven* faces: three vertices, three edges and the simplex itself. The relative interiors of the faces, i.e., vertexes themselves, simplex edges without its boundaries (vertexes) and the interior of the simplex partition Σ.

Lemma 2. *Let (e^*, σ^*) with $\sigma^* \in \tilde{F}^k$. Suppose (e^*, σ^*) is an empathetic equilibrium. Then, there exists $u^* \in \mathbb{R}^{I\times|\mathcal{S}^k|}$ that generates beliefs e^* on the support, and $v(u^*, \sigma^*)$ belongs to $(L^k)^\perp$.*

Proof: Suppose $v(u^*, \sigma^*) \notin (L^k)^\perp$. Then there exists $\delta \in L^k$ so that $v(u^*, \sigma^*) \cdot \delta \neq 0$. Next, since $\tilde{F}^k$ is an open set contained in L^k, there exists $\alpha \in \mathbb{R}$ such that $\sigma^* + \alpha\delta \in \tilde{F}^k$ and $v(u^*, \sigma^*) \cdot \alpha\delta = \sum_{i\in I} v_i(u^*, \sigma^*) \cdot \alpha\delta_i > 0$. So $v_i(u^*, \sigma^*) \cdot \alpha\delta_i > 0$ for some player i. Finally, $\sigma_i^* + \alpha\delta_i \in \Delta(S_i^k)$, and so player i has a profitable deviation, contradicting Definition 1-ii).□

The affine subspace L^k is an H-dimensional smooth manifold where $H \leq \left(\sum_{i\in I} |S_i|\right) - I$. Let $T^k = \{t_h^k\}_{h=1}^H$ be an orthogonal base of L^k. As in Lemma 1, consider an open set of perturbations $\mathcal{P} \subset \mathbb{R}^{I\times|\mathcal{S}|}$. Define a function $f^k : \mathbb{R}^{I\times|\mathcal{S}^k|} \times L^k \times \mathcal{P} \to \mathbb{R}^{I\times|\mathcal{S}^k|} \times \mathbb{R}^H$, where:

$$f^k(u, \sigma, p) \equiv \begin{pmatrix} (\mathcal{U}(s, u_s) - u_s + p_s)_{s\in\mathcal{S}^k}, \\ (v(u, \sigma) \cdot t_h^k)_{h=1}^H \end{pmatrix}$$

For motivation, consider a set of solutions to $f^k(u, \sigma, p) = 0$. Observe that the first $I \times \left|\mathcal{S}^k\right|$ equations fix a utility vector u that generates beliefs on the support $\mathcal{S}^k$, whereas the last H equations are necessary equilibrium conditions (Lemma 2). Altogether, if (u^*, σ^*) is an empathetic equilibrium in a perturbed game Γ_p, then (u^*, σ^*) must be a root of $f^k(\cdot, p)$.

Lemma 3. *Consider Assumption 1. For each face F^k, there exists a subset $\mathcal{P}^k$ of $\mathcal{P}$ with full Lebesque measure such that for all $p \in \mathcal{P}^k$, the set $(f^k)^{-1}(0, p)$ contains only isolated points.*

Proof: We first show that function f^k is transverse to zero ($f^k \pitchfork 0$). First, each of the first $I \times |\mathcal{S}^k|$ components can be independently perturbed by $p = (p_s)_{s\in\mathcal{S}^k}$. Second, note that, for every strategy $s_i \in S_i^k$, the expected payoff $v_i(s_i|u, \sigma) = \sum_{s_{-i}\in\mathcal{S}_{-i}} \prod_{j\neq i} \sigma_j(s_j) u_i(s_i, s_{-i})$ can be made arbitrarily by choosing $u_i(s_i, s_{-i})$ accordingly. Third, for every profile $s \notin \mathcal{S}^k$, each (basis) vector t_h^k is multiplied by zero. Thus, adjusting v in the direction of t_h^k can independently perturb each of the orthogonality conditions. This change in v would affect only the value of the h^{th} condition, leaving all others unchanged. Altogether, the Jacobian of f^k has full rank $I \times \left|\mathcal{S}^k\right| + H$, so f^k is transverse to zero ($f^k \pitchfork 0$).[39]

Next we argue that each root of f^k is isolated. First, by the Transversality Theorem, there exists $\mathcal{P}^k \subset \mathcal{P}$ with full Lebesque measure such that $f^k(\cdot, p) \pitchfork 0$ for all $p \in \mathcal{P}^k$. Thus,

[39] Given σ, the Jacobian of f^k, Df_k, is a block matrix:

$$Df_k = \left(\begin{array}{r|cc} & D_p f^k & D_u f^k \\ \hline \text{consistency } e & \mathbb{I}_{I\times|\mathcal{S}^k|} & D_u(\mathcal{U} - u) \\ \text{orthogonality } v & 0 & (D_u v)T \end{array}\right)$$

where $\mathbb{I}_{I\times|\mathcal{S}^k|}$ is an identity matrix, $D_u v$ is the Jacobian of v in u, and $D_u(\mathcal{U} - u)$ is that of $\mathcal{U} - u$. Since $D_u v$ has full rank, and the vectors in T^k are orthogonal, the product matrix has full rank h. So Df_k has full rank.

for any root $(u,\sigma) \in (f^k)^{-1}(0,p)$, the Jacobian of $f^k(\cdot,p)$ has full rank $I \times |\mathcal{S}^k| + H$. Finally, by the Inverse Function Theorem, there exists a neighborhood of (u,σ), where $f^k(\cdot,p)$ is a bijection, which contains at most one root. Thus, the root (u,σ) is isolated. $\square$

Lemma 4. *Consider Assumption 1. There exists a set $\bar{\mathcal{P}} \subset \mathcal{P}$ with full Lebesque measure such that the equilibrium set in the game Γ_p is finite, for all $p \in \bar{\mathcal{P}}$.*

Proof: First, for any $p \in \mathcal{P}$ and $\mathcal{U}$ satisfying Assumption 1, one can find a closed box $B_u \subset \mathbb{R}^{I \times |\mathcal{S}|}$ so that for any $u = (u_s)_{s \in \mathcal{S}} \in B_u$, the utility vector u_s obeys $\mathcal{U}(s, u_s) + p_s = u_s$. Second, let $B_\Sigma \subset \mathbb{R}^{\sum_i |S_i|}$ be a closed box containing Σ, and for each face F^k, let $B^k_\Sigma \equiv B_\Sigma \cap L^k$. Also, consider the set $B_u \times B^k_\Sigma$. By construction, all profiles (u^*, σ^*) with $\sigma^* \in \tilde{F}^k$ that can generate equilibria belong to $B_u \times B^k_\Sigma \cap \left(f^k\right)^{-1}(0,p)$. Since $B_u \times B^k_\Sigma$ is compact and $\left(f^k\right)^{-1}(0,p)$ contains only isolated points (Lemma 3), the set $B_u \times B^k_\Sigma \cap \left(f^k\right)^{-1}(0,p)$ is finite.[40] Next, by Lemma 3, there exists $\{\mathcal{P}^k\}_{k=1}^K$ such that $(f^k)^{-1}(0,p)$ contains only isolated points for all $p \in \mathcal{P}^k$. Consider $p \in \bar{\mathcal{P}} \equiv \bigcap_{k=0}^K \mathcal{P}^k$ (the union includes set $\mathcal{P}^0$ defined in Lemma 1). Since $B_u \times B^k_\Sigma \cap \left(f^k\right)^{-1}(0,p)$ is finite for each k, it follows that $\bigcup_{k=1}^K (B_u \times B^k_\Sigma) \cap \left(f^k\right)^{-1}(0,p)$ is finite too. Also, since $\{\tilde{F}^k\}_{k=1}^K$ is a partition of Σ, we have that for every $\tilde{F}^k \subset B^k_\Sigma$:

$$B_u \times \Sigma \cap \left(f^k\right)^{-1}(0,p) \subset \bigcup_{k=1}^K (B_u \times B^k_\Sigma) \cap \left(f^k\right)^{-1}(0,p)$$

Thus, the set $B_u \times \Sigma \cap \left(f^k\right)^{-1}(0,p)$ is finite. Note that each $(u^*, \sigma^*) \in B_u \times \Sigma \cap \left(f^k\right)^{-1}(0,p)$ can generate at most one strategy profile with "support" beliefs. Finally, the set of consistent beliefs for profiles s not in the support is finite by that fact that, for $p \in \mathcal{P}^0$, the utility set $\mathbf{U}(\cdot)$ is finite for any strategy (Lemma 1). Altogether, the set of mixed empathetic equilibria is finite, for all $p \in \bar{\mathcal{P}}$. $\square$

We can now conclude the proof of Proposition 1. First, the utility set $\mathbf{U}(\cdot)$ is finite in any generic game, by Lemma 1. Second, by Lemma 2–3, the set of equilibria necessarily contains only isolated points. Next, by Lemma 4, there is a finite number of such points. Finally, since each equilibrium (e^*, σ^*) generates one outcome o^*, the equilibrium set is generically finite, because $\mathcal{O}^*$ is generically finite. $\square$

Proof of Proposition 2: First, any empathetic equilibrium is a weak empathetic equilibrium, thus $\mathcal{O}^* \subseteq \mathcal{O}^{**}$. Conversely, fix an equilibrium outcome $o^* = (\sigma^*, v^*) \in \mathcal{O}^{**}$ induced by some weakly consistent empathetic beliefs e^{**}. Next, we construct consistent beliefs e^* that also induce equilibrium outcome o^*. For the sake of clarity, now we introduce notation needed only here. For any $i \in I$, the deviation set of player i is $S^d_i =$

[40] See the proof of Lemma 1 for a more elaborated argument.

$\{s \in \mathcal{S}_0 : s_i \neq \text{supp}(\sigma_i^*) \text{ and } s_j \in \text{supp}(\sigma_j^*) \text{ for all } j \neq i\}$. The residual set contains all profiles that are neither in a deviation set nor in the support of σ^*, that is, $\mathcal{S}^r \equiv \mathcal{S} \backslash \left(\bigcup_{i \in I} S_i^d \cup \text{supp}(\sigma^*)\right)$. Now consider the following beliefs e^*. For all profiles $s \in \text{supp}(\sigma^*)$, beliefs are as in the original equilibrium: $e_i^*(s) \equiv e_i^{**}(s)$ for all $i \in I$. Next, for any $i \in I$ and any profile $s \in S_i^d$, beliefs $e^{**}(s)$ are weakly consistent, and so there must exist a utility profile $u(s) \equiv (u_i(s), e_i^{**}(s)) \in \mathbf{U}(s)$. Let $e_i^*(s) \equiv e_i^{**}(s)$ for player i and $e_j^*(s) \equiv u_{-j}(s)$ for all $j \neq i$. Next, for any $s \in \mathcal{S}^r$, pick any solution $u(s) \in \mathbf{U}(s)$ and define $e_i^*(s) \equiv u_{-i}(s)$ for all $i \in I$ (at least one such profile exists since o^* is a weak equilibrium outcome). Since the sets $\text{supp}(\sigma^*), \mathcal{S}^r$ and $\{S_i^d\}_{i \in I}$ form a partition of the strategy space $\mathcal{S}$, beliefs e^* are defined on the entire domain $\mathcal{S}$. Also, by construction, beliefs e^* are consistent at all $s \in \mathcal{S}$.

Now we argue that σ^* is a mutual best response given e^*. Fix a player i. For any $\sigma_i \in \Delta(S_i)$, we have $\text{supp}(\sigma_i, \sigma_{-i}^*) \subset \text{supp}(\sigma^*) \cup S_i^d$. Since on these sets i's beliefs are unchanged, $e_i^*(s) = e_i^{**}(s)$ we have $U_i^R(\sigma_i, \sigma_{-i}^* | e_i^*) = U_i^R(\sigma_i, \sigma_{-i}^* | e_i^{**})$ by (1). This means that if σ_i^* is a best response to σ_{-i}^* given beliefs e_i^{**}, then it is a best response given e_i^*. This logic holds for all players $i \in I$. So we have that beliefs e^* are consistent at any $s \in \mathcal{S}$, and σ^* is a Nash equilibrium given e_i^*. Finally, for each $i \in I$, $v_i^* = U_i^R(\sigma^* | e_i^*) = U_i^R(\sigma^* | e_i^{**})$, and so equilibrium (σ^*, e^{**}) induces outcome $o^* = (\sigma^*, v^*)$. Altogether, $\mathcal{O}^{**} \subseteq \mathcal{O}^*$. $\square$

Let the set of all distinct selections of $\mathbf{U}(\cdot)$ be denoted by $\mathcal{R}$ and let u^r be its typical element.[41] Call $\mathcal{O}^{*r}$ the set of equilibrium outcomes, given u^r.

Claim A.1.1 (Decomposition). *Suppose Assumption 1 holds. The set of equilibrium outcomes satisfies* $\mathcal{O}^* = \bigcup_{r \in \mathcal{R}} \mathcal{O}^{*r}$.

Proof: First, as in the proof of Proposition 1, Assumption 1 implies that utility set is nonempty, $\mathbf{U}(s) \neq \emptyset$ for any $s \in \mathcal{S}$. Next, consider an equilibrium outcome $o^* = (\sigma^*, v^*) \in \mathcal{O}^*$. By definition of equilibrium there exist consistent beliefs e^*, such that σ^* is Nash equilibrium given e^* and (σ^*, e^*) induce utilities v^*. Now for every profile s and player i, let $u_i^r(s) \equiv U_i(s, e_i^*(s))$. Then, σ^* is a Nash equilibrium in $\Gamma^r = \langle I, \mathcal{S}, u^r \rangle$, and so $o^* \in \mathcal{O}^{*r} \subseteq \bigcup_{r \in \mathcal{R}} \mathcal{O}^{*r}$. Conversely, consider $o^* = (\sigma^*, v^*) \in \bigcup_{r \in \mathcal{R}} \mathcal{O}^{*r}$. Then σ^* is a Nash equilibrium for some reduced-form utilities $u^r(\cdot) \in \mathbf{U}(\cdot)$. Next, for every (i, s), let $e_i^* \equiv u_{-i}^r(s)$. Since $U_i^R(s | e_i^*) = u_i^r(s)$ for all s and i, the pair (σ^*, e^*) is an equilibrium, and so $\bigcup_{r \in \mathcal{R}} \mathcal{O}^{*r} \subseteq \mathcal{O}^{**}$. $\square$

Example 3.2 has two reduced-form games ($\mathcal{R} = \{1, 2\}$):

	G	C
G	$0, 0$	$-1, 1$
C	$1, -1$	$1, 1$

	G	C
G	$2, 2$	$-1, 1$
C	$1, -1$	$1, 1$

[41] Observe that if $\mathbf{U}(\cdot)$ is finite, there are $|\mathcal{R}| \equiv \prod_{s \in \mathcal{S}} |\mathbf{U}(s)| < \infty$ distinct reduced-form games.

In the left reduced-form game, the set of equilibrium outcomes is a singleton: $\mathcal{O}^{*1} = \{[(C,C),(1,1)]\}$. In the right game, the equilibrium outcome set has three elements: $\mathcal{O}^{*2} = \{[(C,C),(1,1)],[(G,G),(2,2)],[(2/3,1/3)_{i=1,2},(1,1)]\}$, where 2/3 is the chance of playing G. Notice that the outcome of the left game is also an element of the outcome set of the right game; thus the equilibrium outcome set coincides with $\mathcal{O}^{*2}$.

Proposition A.1.1. *The set of Pareto efficient outcomes $\mathcal{O}^{Pareto}$ is non-empty.*

Proof: Since $\mathcal{U}$ is continuous and bounded (Assumption 1), we have that for any profile $s \in \mathcal{S}$, the set $\mathbf{U}(s)$ is non-empty, closed, and bounded. Thus, the set $X_0 \equiv \Sigma \times (\times_{s\in\mathcal{S}} \mathbf{U}(s))$ is non-empty and compact. Next, for any $i \in I$ define sets $X_i, i \in I$, recursively

$$X_i = \arg \max_{(\sigma,u)\in X_{i-1}} \sum_{s\in\mathcal{S}} \prod_{j\in I} \sigma_j(s_j) u_i(s)$$

Note that $X_i \subseteq X_{i-1}$; also, $X_i \neq \emptyset$ and compact, for $X_{i-1} \neq \emptyset$ and compact (Maximum Theorem). Next, fix a tuple $(\sigma,u) \in X_I$. We will show that (σ,u) induces a Pareto efficient outcome o. Consider any $o' \in \mathcal{O}$ characterized by $(\sigma',u') \in X_0$. If $(\sigma',u') \in X_I$, then by definition, $(\sigma,u),(\sigma',u') \in X_i$ for all i, and so (σ',u') does not Pareto dominate (σ,u). Conversely, if $(\sigma',u') \in X_0/X_I$, then there exists i such that $(\sigma',u') \in X_{i-1}$ but $(\sigma',u') \notin X_i$. Since X_i is the set of all profiles that maximize player's i payoff on X_{i-1}, (σ,u) must leave player i better off compared to (σ',u'). Thus, (σ',u') does not Pareto dominate (σ,u); therefore, any outcome o characterized by $(\sigma,u) \in X_I$ is Pareto efficient. Finally, $\mathcal{O}^{Pareto}$ is non-empty, because the set X_I is non-empty, $X_I \neq \emptyset$. □

Example A.1.1. *Consider a version of Example 3.2:*

	G	C
G	$\sqrt{2u_b}, \sqrt{2u_a}$	$-1,-1$
C	$-1,-1$	$-2,-2$

Here strategy G is strictly dominant for all consistent beliefs. So this game has two equilibrium outcomes: $o^*_H = ((G,G),(2,2))$ and $o^*_L = ((G,G),(0,0))$. Clearly, o^*_H Pareto dominates o^*_L. In this example, players coordinate beliefs on an inefficient solution to the interdependent utility system. In other words, Inefficiencies do not arise because of the strategies that both agents elect, but because of the payoff level at which they coordinate their beliefs, given their choice (G,G).

A.2 Proofs of Section §5

Proof of Proposition 3: First, by Assumption 1, the utility set $\mathbf{U}(s) \neq \emptyset$ for all $s \in \mathcal{S}$. Also, since utility functions $(U_i)_{i\in I}$ are continuous and bounded, we have that for any profile s, the set $\mathbf{U}(s)$ is non-empty, closed, and bounded. Thus, $\mathbf{U}_i(s)$ is compact, and so $\inf \mathbf{U}_i(s) = \min \mathbf{U}_i(s)$. Next, consider an outcome (σ, v) that violates the best response condition, so that $\underline{v}_i(\sigma_{-i}) > v_i$ for some i. Take any consistent beliefs $e(\cdot)$, satisfying $v_i = \sum_{s\in\mathcal{S}} U_i(s, e_i(s))\sigma(s)$ for all i. Because $\underline{v}_i(\sigma_{-i}) > v_i$ for some i, there must exist a strategy $s_i' \neq s_i$ such that:

$$v_i < \sum_{s_{-i}\in\mathcal{S}_{-i}} \min \mathbf{U}_i(s_i', s_{-i})\sigma_{-i}(s_{-i}) \leq \sum_{s_{-i}\in\mathcal{S}_{-i}} U_i(s_i', s_{-i}, e(s_i', s_{-i}))\sigma_{-i}(s_{-i}),$$

where the last equality follows by the definition of $\mathbf{U}_i$ and the fact that beliefs e are consistent. Thus, s_i' is a profitable deviation for player i, and so (s, v) cannot be an equilibrium outcome.

Conversely, fix an outcome (σ^*, v^*) with $\underline{v}_i\left(\sigma^*_{-i}\right) \leq v_i^*$ for all i (and so $\underline{v}_i < \infty$). We introduce some notation to make the argument clearer. For any $i \in I$, define i's deviation set as $\mathcal{S}_i^d \equiv \{(s_i, s^*_{-i}) \in \mathcal{S} : s_i \notin \operatorname{supp}(\sigma_i^*) \text{ and } s^*_{-i} \in \operatorname{supp}(\sigma^*_{-i})\}$ and the residual set as $\mathcal{S}^r \equiv \mathcal{S}/\left(\bigcup_{i\in I}\mathcal{S}_i^d \cup \operatorname{supp}(\sigma^*)\right)$. By construction, $\operatorname{supp}(\sigma^*), \left\{\mathcal{S}_i^d\right\}_{i\in I}, \mathcal{S}^r$ partition the strategy profile space $\mathcal{S}$. Now we define weakly consistent beliefs $e^*(\cdot)$. First, for each $s^* \in \operatorname{supp}(\sigma^*)$, take $u(s^*) \in \mathbf{U}(s^*)$ ensuring that $v^* = \sum_{s^*\in\operatorname{supp}(\sigma^*)} u(s^*)\sigma^*(s^*)$ and $v_i^* = \sum_{s^*_{-i}\in\operatorname{supp}(\sigma^*_{-i})} u(s_i^*, s^*_{-i})\sigma^*_{-i}(s^*_{-i})$ for all $s_i^* \in \operatorname{supp}(\sigma_i^*)$, and for all player $i \in I$. Then, assign $e_i^*(s^*) = u^*_{-i}(s^*)$ for all $i \in I$. Next, for any i and $s \in \mathcal{S}_i^d$, take $u(s) \in \mathbf{U}(s)$ such that $u_i(s) = \min \mathbf{U}_i(s)$, and for all players let $e_j^*(s) = u_{-j}(s)$. Finally, for $s \in \mathcal{S}^r$, take any $u(s) \in \mathbf{U}(s)$ and let $e_i^*(s) = u_{-i}(s)$ for all i. Altogether, e^* is weakly consistent for all $s \in \mathcal{S}$.

Next, by construction of beliefs, it is enough to assure that no player has incentives to deviate to a pure strategy not in the support of σ^*. This is indeed the case, for since $U_i(s, e_i^*(s)) = \min \mathbf{U}_i(s_i, s^*_{-i})$ for all $s \in \mathcal{S}_i^d$ and $v_i^* \geq \underline{v}_i\left(\sigma^*_{-i}\right)$, it follows that:

$$v_i^* \geq \sum_{s^*_{-i}\in\operatorname{supp}(\sigma^*_{-i})} \min \mathbf{U}_i(s_i, s^*_{-i})\sigma^*_{-i}(s^*_{-i}) = \sum_{s^*_{-i}\in\operatorname{supp}(\sigma^*_{-i})} U_i(s_i, s^*_{-i}, e_i(s_i, s^*_{-i}))\sigma^*_{-i}(s^*_{-i}),$$

for all $s_i \notin \operatorname{supp}(\sigma_i^*)$. Finally, since this logic holds for all i, (s^*, e^*) is a weak equilibrium inducing a utility profile v^*. Thus, (s^*, v^*) is an equilibrium outcome, by Proposition 2. □

A.3 Proofs of Section §6

Claim A.3.1. *For any profile $s \in \mathcal{S}$ there exist a reduced form utility profile $\underline{u}_s, \overline{u}_s \in \mathbf{U}(s)$ such that $\underline{u}_s \leq u_s \leq \overline{u}_s$ for all $u_s \in \mathbf{U}(s)$. Also, in games with two players, the set $\mathbf{U}(s)$ is totally ordered: for any $u_s, u_s' \in \mathbf{U}(s)$ and $i \neq j$ we have $u_{i,s} \geq u'_{i,s}$ iff $u_{j,s} \geq u'_{j,s}$.*

Proof: Fix $s \in \mathcal{S}$. Since the utility system $\mathcal{U}(s, \cdot)$ is bounded (Assumption 1), there exists a closed box $B \subset \mathbb{R}^I$ such that the image of $\mathcal{U}$ (and hence all its fixed points) is in B. Consider a restriction of $\mathcal{U}(s, \cdot)$ to B. Since $\mathcal{U}(s, \cdot)$ is increasing and B is a complete lattice (it is a closed and bounded box), the set $\mathbf{U}(s)$ is a non-empty complete lattice (so it has a maximal and minimal element), by Tarski's Fixed Point Theorem. Next, consider a two-player game, and let $u_s, u'_s \in \mathbf{U}(s)$ with $u_{i,s} \geq u'_{i,s}$. Thus, we have $u_{j,s} = U_j(s, u_{i,s})$ and $u'_{j,s} = U_j(s, u'_{i,s})$, for $j \neq i$. Since $U_j(s, \cdot)$ is increasing, we have $u_{j,s} = U_j(s, u_{i,s}) \geq U_j(s, u'_{i,s}) = u'_{j,s}$. Reversing the roles of i and j yields $u_{j,s} \geq u'_{j,s}$ iff $u_{i,s} \geq u'_{i,s}$. $\square$

Claim A.3.2. *Let $I = 2$ and fix a profile s. The utility set $\mathbf{U}(s)$ has, at most, two elements.*

Proof: Suppose wlog that both players exhibit diminishing empathy. For $i = 1, 2$ let $Y_i \equiv U_i(s, \mathbb{R})$ be a target set of utility function $U_i(s, \cdot)$. Since $U_i(s, \cdot)$ is strictly concave and hence continuous, by the Intermediate Function Theorem Y_i is convex. Let $\tilde{U}_i : Y_j \to Y_i$ be a restriction of $U_i(s, \cdot)$ to Y_j. Observe that any $u \in \mathbf{U}(s)$ necessarily satisfies $u \in Y_1 \times Y_2$, and hence it is a solution to $\tilde{U}_i(u_j) = u_i$ for $i = 1, 2$. By construction $\tilde{U}_i$ is surjective and since $U_i(s, \cdot)$ is increasing and strictly concave, it is also strictly increasing and hence $\tilde{U}_i$ is injective. It follows that inverse function $\tilde{U}_2^{-1} : Y_1 \to Y_2$ is well defined. Finally, $\tilde{U}_2$ is increasing and strictly concave, and hence, inverse $\tilde{U}_2^{-1}$ is increasing and strictly convex.

Let $\varphi : Y_1 \to Y_2$ be defined as $\varphi(x) = \tilde{U}_1(x) - \tilde{U}_2^{-1}(x)$. Observe that a vector $(u_1, u_2) \in \mathbf{U}(s)$ iff $\varphi(u_1) = 0$ and $u_2 = \tilde{U}_2(u_1)$. Function φ is the sum of two increasing strictly concave functions, namely, $\tilde{U}_1$ and $-\tilde{U}_2^{-1}$, hence it is strictly concave, and as such it can have at most two roots. Otherwise one could find $u'_1 > u''_1 > u'''_1$ in Y_1 such that $\varphi(u'_1) = \varphi(u''_1) = \varphi(u'''_1)$, which contradicts the strict concavity of $\varphi(\cdot)$. $\square$

Claim A.3.3. *Fix $s \in \mathcal{S}$ with $\mathcal{U}$ satisfying Inada. Generically, $\mathbf{U}(s)$ has two elements or none.*

Proof: Let $\tilde{U}_2^{-1}$ as in the proof of Claim A.3.2. The limit $\lim_{u_1 \downarrow \inf Y_1} \partial \tilde{U}_2^{-1} / \partial u_1 = 0$ and $\lim_{u_1 \uparrow \sup Y_1} \partial \tilde{U}_2^{-1} / \partial u_1 = \infty$, since U_2 satisfies Inada. Next, we claim that if $U_2(s, u_1) = \tilde{U}_2^{-1}$ and $\partial U_2(s, u_1)/\partial u_1 \neq \partial \tilde{U}_2^{-1}/\partial u_1$, then there must exist $u_1^* \neq u_1$ such that $U_2(s, u_1^*) = \tilde{U}_2^{-1}(s, u_1^*)$. Suppose wlog that $\partial U_2(s, u_1)/\partial u_1 > \partial \tilde{U}_2^{-1}(s, u_1)/\partial u_1$. Then there exists a small $\varepsilon > 0$ such that $U_2(s, u_1 + \varepsilon) > \tilde{U}_2^{-1}(s, u_1 + \varepsilon)$. But, since the respective slopes of U_2 and $\tilde{U}_2^{-1}$ vanish and explode as $u_1 \uparrow \infty$, there exists a large $\eta > 0$ such that $U_2(s, u_1 + \eta) < \tilde{U}_2^{-1}(s, u_1 + \eta)$. But then, by the Intermediate Value Theorem, there must exist $u_1^* \in (u_1 + \varepsilon, u_1 + \eta)$ with $U_2(s, u_1^*) = \tilde{U}_2^{-1}(s, u_1^*)$. Altogether, if there is unique u_1 with $U_2(s, u_1) = \tilde{U}_2^{-1}(s, u_1)$, then $\partial U_2(s, u_1)/\partial u_1 = \partial \tilde{U}_2^{-1}(s, u_1)/\partial u_1$. Since $u_2 = \tilde{U}_2^{-1}(s, u_1)$, a unique fixed point implies that the Jacobian of $\mathcal{U}(s, u) - u$ is singular at $u = (u_1, u_2)$. But then considering a perturbed utility system $\mathcal{U}(s, u) + p$ with $p \in \mathcal{P} \subset \mathbb{R}^I$, and by the same logic of the proof of Lemma 1, the Jacobian of $\mathcal{U}(s,u) - u$ is singular only for a negligible set of perturbations. Finally, by Lemma A.3.2, the set $\mathcal{U}(s)$ has generically either none or two fixed points. $\square$

Claim A.3.4. *Fix $s \in \mathcal{S}$. For any $u_s, u'_s \in \mathbf{U}(s)$ and $i \neq j$ we have: $u_{i,s} \geq u'_{i,s}$ iff $u_{j,s} \leq u'_{j,s}$.*

Proof: Let $u_s, u'_s \in \mathbf{U}(s)$ with $u_{i,s} \geq u'_{i,s}$. Thus, since $U_j(s,\cdot)$ is decreasing for $j \neq i$, we have $u_{j,s} = U_j(s, u_{i,s}) \leq U_j = u'_{j,s}$. Reversing the roles of (i,j) yields $u_{j,s} \geq u'_{j,s}$ iff $u_{i,s} \leq u'_{i,s}$. $\square$

Observation A.3.1. *The existence of a symmetric solution is a general feature of symmetric empathic games. An empathic game is* symmetric *if $S_i = S_j$ and $U_i(s,\cdot) \equiv U_j(s,\cdot)$ for all players $i \in I$. Under mild conditions,* a symmetric reduced-form utility profile always exists. *Suppose that $u_i = x \in \mathbb{R}$ for all $i \in I$, and let $\hat{U}(s,\cdot) : \mathbb{R} \to \mathbb{R}$ where $\hat{U}(s,x) \equiv U(s,u_{-i})$. Clearly, a symmetric reduced form utility profile is a fixed point of $\hat{U}(s,x)$. Assume that $U(s,\cdot)$ is differentiable. Then, by the antipathy assumption: $(d/dx)\hat{U}(s,x) = \sum_{\ell \neq i}(\partial/\partial u_\ell)U(s,u_\ell)|_{u_\ell = x} < 0$. Thus, by the Intermediate Value Theorem, a symmetric payoff vector exists iff there exists $\underline{x}, \bar{x} \in \mathbb{R}$ with $\hat{U}(s,\underline{x}) \leq 0 \leq \hat{U}(s,\bar{x})$. In the example of §6.2, the utility function $U(s,\cdot)$ is continuous and obeys $U(s,0) = 1 \geq U(s,1) = 0$ at $s = (G,G)$.*