THE UNIVERSITY OF CHICAGO

PAYMENT BETWEEN FRIENDS: COMPENSATION AND SIGNALING IN AN ALTRUISTIC PARTNERSHIP

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY
SETH HARRIS BLUMBERG

CHICAGO, ILLINOIS
DECEMBER, 2019

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Friends often help each other without direct compensation. This behavior seems to contravene a basic tenet of economics, that prices and payment help coordinate supply and demand, increasing welfare. If compensation is useful between strangers, why not between friends as well?

I address this question by building a model of bilateral trade in a friendship, where the price must be specified ex ante, and use it to understand under which circumstances a price of zero is an equilibrium outcome. In the model, two people have altruistic preferences towards each other, but may also have incomplete information about each other's motives. While altruists are able to undertake some beneficial transactions even at a price of zero, they would still benefit from a more efficient price. In contrast to previous work, I find that choosing to forgo payment cannot arise from being altruistic, nor from wanting to signal one's altruism. The key driver of this result is that the benefit of an altruistic reputation is greater to a selfish person than to an altruistic one, so no separating equilibrium in a signaling game could be sustained in which only the altruistic type sets price to zero.

Instead, I find that a separating equilibrium is possible where a price of zero signals a partners *trust* in the others altruism, though this result can only be sustained at extreme parameters. Overall, my model rules out altruism signaling as a fundamental reason for the lack of priced transactions between friends, but opens the possibility that this behavior could be driven by signaling of trust. I conclude by discussing alternative approaches to examining this widespread behavior.

INTRODUCTION

Motivation and Overview

As every student of economics learns, money is useful to facilitate trade. It incentivizes producers to produce when their cost is low, and consumers to consume when their benefit is high, coordinating supply and demand. In models of exchange, transfers constitute an essential piece of how value is created and traded.

And yet, plenty of value is created without payment. People willingly do favors for their friends and colleagues all the time, seemingly for free. When we receive goods and services from those close to us, we seldom give them direct compensation. Evidently, some alternative motives are what drive us to help each other out.

Still, if money is so useful between strangers, why do we refrain from using it between friends? This behavior is a puzzle, because the logic behind the usefulness of money does not require that the parties involved be selfish. Is there some reason why payment cannot be combined with those alternative motives?

Motivated by these questions, in this paper I build a model of bilateral trade in a friendship. In the model, two friends feel altruism for one another, but they may also be uncertain about each other's motives or beliefs.[1] The model has two periods. In period 1, at the outset of their friendship, they set the ground rules, deciding what price p—possibly zero—they will charge each other when one performs a service for the other. In period 2, when an opportunity does arise where one needs the other's help, they play a bilateral bargaining game in which one (the "buyer") decides whether to request the service, and the other (the "seller") decides whether to provide it, at that price p. In period 1, they do not yet know their roles (buyer and seller), and in period 2, their valuations (cost and benefit) are private

1. The model could also represent any relationship with mutual help and altruism, such as between colleagues, family members, or romantic partners.

information.

I use the model to study the use (or non-use) of compensation between friends. I pose and answer two related sets of questions. First, when used between friends, does money retain or lose its usefulness? How does this usefulness depend on the altruism preferences of the two friends towards each other? Second, when would someone choose to set zero as the price for services between them and their friend? Under what conditions can such a choice constitute an equilibrium outcome, and if it emerges as a separating equilibrium, what might the actions the a zero or nonzero price signal?

Chapter 1 answers the first question, on the usefulness of compensation between friends. I study the period-1 bilateral trading game, with altruism preferences as common knowledge. I define the ***marginal value of compensation*** in the friendship as the slope of the value of the partnership with respect to price, at $p = 0$. When it is negative to an individual i, she (locally) prefers a price of zero over a slightly higher price; in a unit-uniform specifications, it also indicates she (globally) prefers price of zero over other prices. I then conduct comparative statics on the marginal value of compensation with respect to the two parties' altruism. The main results are that, for an individual i, her marginal value of compensation is increasing in her own altruism α_i and decreasing in her friend j's altruism α_j; this means that compensation is a complement to α_i and a substitute to α_j, in producing i's utility. Further, if α_i is sufficiently low and α_j is sufficiently high, then the latter effect may dominate, making the marginal value of compensation negative to i. In this case, i would prefer a price of zero.

Chapter 2 extends Chapter 1's bilateral trading model to include uncertainty by each player about her partner's preferences and beliefs. This allows me to answer the second question, which asks what circumstances would allow an equilibrium where the friends choose zero as the price for services between them. I explicitly model a two-period game, where in period 1, one of the two chooses a price, and then in period 2, they play the bilateral trading game

of Chapter 1.

Our analysis focuses on understanding when there exists Perfect Bayesian Equilibrium of the period-1 game in which a price of zero is chosen. There are three main results, based on three information environments:

- First, when altruism preferences are common knowledge, then an individual i's equilibrium choice of price cannot be zero unless (following Chapter 1's result) she is sufficiently selfish relative to her friend j, so that the marginal value of compensation is negative to her. This does not seem a compelling model of a healthy friendship, but rather like i exploiting j's kindness.

- Second, when j is uncertain about i's altruism, there *cannot* exist a separating equilibrium where the more altruistic type of i chooses $p = 0$ while the more selfish type of i chooses a $p > 0$. In any situation where the altruistic type of i would choose $p = 0$, the selfish type would choose $p = 0$ as well. That is, choosing to forgo cash cannot be a signal of altruism.

- Third, when j has uncertainty about i's altruism, and i also has (second-order) uncertainty about j's beliefs, there may exist a separating equilibrium in which $p = 0$ is chosen by the more trusting type of j (who believes that i is likely altruistic), while a higher price is chosen by the less trusting type of j (who believes that i is likely selfish). The logic is that when j convinces i that she is trusting, then both i and j behave closer to what they wish they could commit to – depending on the price, either more generously (seller choosing to sell even at high cost, buyer asking selectively only for high-value services), or less generously. This is an outcome they both would like to be able to commit to, relative to equilibrium. For this separating equilibrium to exist, both the altruism of the high type of i and the belief of the optimistic type of j must be very high, so that the benefits of coordinating outweigh the classical incentive

benefits of a higher price.

Together, these three results imply that if someone chooses $p = 0$ within the friendship, either she is taking advantage of her friend's kindness, or she is signaling that she has trust in her friend's altruism; it cannot be that she is signaling her own altruism.

Naturally, the model is not without its limitations. First, it employs some specific assumptions for the sake of tractability. For some results, it assumes that valuations (cost and benefit) are drawn independently from the unit uniform distribution ($\texttt{Unif}[0, 1]$). While this makes equations easier to solve, it also makes the results more particular. The model also assumes that if trade occurs, it does so at a single pre-agreed price of p, regardless of the valuation draws. While this behavior is known to be an inefficient equilibrium in this bilateral trading game (Chatterjee and Samuelson, 1983), it does reduce the period-1 choice to a single number, rather to a more general mechanism

Second, the model restricts the choice of price p to the interval $[0, 1]$, the same as the support of the valuations for the service. In particular, this rules out negative prices, where the "seller" actually pays the "buyer" to let her provide the service. While this simplifies the analysis, it does affect the interpretation of certain result; when the analysis indicates that a price of zero is preferred over higher prices, it may not in fact be preferred over negative prices.

Third, the model shuts of dynamics. A plausible explanation for what friends are doing when they help each other "for free" is that they in fact will pay each other back later, by returning a favor. However, even if we accept such behavior, the original question remains: why trade favors, rather than pay back immediately? Section 3.1 in Chapter 3 discusses how one might extend this paper's question to comparing a wider range of mechanisms, including a dynamic favor-trading mechanism, such as building off of Möbius (2001) or Abdulkadiroğlu and Bagwell (2013). This paper does not go down that route, because

incorporating reputation dynamics into this repeated game adds additional complications to solving the model and deriving insight from it. The incentives for seeking a reputation would be blunted in such a model, since players would learn each other's types in any case by observing their behavior in early rounds of the game.

Fourth, Chapter 2 studies only Perfect Bayesian Equilibrium, an equilibrium concept that does not restrict the posterior beliefs that follow an out-of-equilibrium choice of p. This freedom is a strength when showing the nonexistence of separating equilibrium for signaling altruism. But, it is a weakness for showing that a trust-signaling equilibrium exists, because it is just one equilibrium among many. In terms of interpreting the result, this means that equilibrium existence should not be a strong prediction that $p = 0$ would be chosen in actual play of this game. Rather, it provides an equilibrium interpretation of this behavior, should it arise.

These limitations are not necessarily insuperable. Indeed, in Section 3.1 I discuss how extensions to the model might address them, and which of the results are likely to survive such an extension. Potentially they could be addressed by a more general model, and some of the key results might still hold.

This paper is not the first seeking to explain why people forgo the use of money with those close to them, nor is it the first to investigate signaling as an explanation. However, it makes two contributions to this literature.

The first contribution is to treat the preference for reputation as endogenous. In most other papers that use signaling to explain the non-use of payment, the model *assumes* that the individual prefers to be seen as pro-social, or knowledgeable about her partner's preferences, or something else. Examples include Camerer (1988), Prendergast and Stole (2001a), Bénabou and Tirole (2006) and Ellingsen and Johannesson (2011).[2] In contrast, this paper provides

2. An exception is Bénabou and Tirole (2003), in which a principal cares about an agent's beliefs only

a microfoundation for why someone might want to appear altruistic (or as trusting in her friend's altruism), based on strategic concerns (i.e. instrumental motives) in a bilateral trading game. By taking this approach, I am able to highlight that a reputation as altruistic may be undesirable, and that even when desirable, the selfish type has more to gain from seeming altruistic than does the altruistic type. This pattern highlights that models assuming that preference for reputation is positively correlated with reputation—or even uncorrelated—is not necessarily a neutral assumption.

The second contribution is to pose the question differently from other papers, by studying a symmetric friendship in which each friend may find herself either on the giving or the receiving end of help. In most other models, at the time when the compensation scheme is chosen, the gift or favor opportunity has already arisen and everybody knows which side of the transaction is theirs. Instead, this model takes an *ex ante* perspective, in which the two decide at the outset what sort of relationship they will have.[3] Without this modeling choice, the seller will have a natural tendency to choose a high price and the buyer to choose a low one. This perspective captures the fact that, even at the start of a partnership, the participants understand what the ground rules will be, namely if they will be paying each other or helping each other out for free.

---

because these beliefs affect her actions.

3. These rules can be interpreted as a choice of equilibrium in the period-1 bilateral trading game. See Section 3.1 for alternate interpretations.

Literature

What role does money play in friendships, and what role *should* it play? This is an old question that continues to arouse debate.

As far as Adam Smith was concerned, trade and friendship were perfectly compatible. "Commerce, ... ought naturally to be, among nations, as among individuals, a bond of union and friendship," he wrote (Smith, 1981 (1776), Vol I, Book IV, Ch III). Still, he did appreciate that people with goodwill for one another do help each other out without trucking and bartering. He reserved his paean to trade to situations where such emotions are absent, famously declaring, "It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love" (Smith, 1981 (1776), Vol I, Book I, Ch II).

On the other side, there is "a general feeling that morally binding relations, especially kinship relations, should be kept apart as far as possible from money transactions" in European and American culture (Bloch, 1989). This attitude, and arguments to justify it, stretches back at least to Aristotle, through Thomas Aquinas to Marx and beyond, although it does differ across cultures (Bloch and Parry, 1989; Åkesson, 2011). The philosopher Michael Sandel, a modern proponent of this view, puts it this way: "To monetize all forms of giving among friends can corrupt friendship" (Sandel, 2012). He argues that when we allocate things using markets and price, we suffuse the interaction with utilitarian norms, and as a consequence, other beneficial aspects of friendship suffer, like the "formative, educative" experience that a true friendship should be.

Neoclassical economics, meanwhile, emphasizes the efficiency of using a price system to allocate goods and services. From such a perspective, the non-use of money or compensation is a puzzle, since it reduces welfare.

Against this neoclassical doctrine, modern economists who engage with this issue ask two questions. First, could it be that money or direct compensation is harmful, and if so, when and why? Second, in situations where people do avoid using money, why? These two questions are of course related – if using money makes people worse off, then they would be wise to avoid it. But the two may yield different answers: perhaps money is not actually harmful, but for some reason people avoid it anyway, making themselves worse off.

One response by economists is simply to champion the neoclassical view, and to polemicize money-avoidance. Gift-giving behavior serves as a prime object of study here, because the classic theory makes such a clear prediction and prescription. Consumer theory holds that people are better off receiving money than a specific gift of equivalent cost (i.e. an in-kind transfer); this claim applies to compensation for work, government transfers, and gifts as well. The logic is straightforward: cash gives the recipient freedom of choice – she could still buy that item, but she could also put the money to any other use, if she preferred.

To test the putative inefficiency of in-kind gifts, Waldfogel (1993) surveyed undergraduates about the Christmas gifts they received. They said they valued these gifts at only 87% of their costs, and so he concluded that a gift of money would have been more efficient. He thus quantified this difference, and unforgettably coined it the "deadweight loss of Christmas". His paper, while influential, was not definitive. Using a different sample (adults at train stations, and faculty, staff, and grad students at Harvard), Solnick and Hemenway (1996) instead found that people valued their Christmas gifts at 214% of cost. Using a more incentive-compatible elicitation methods (an auction to buy back Christmas presents), List and Shogren (1998) also found that respondents valued their Christmas gifts above cost, making them efficient, when counting both a "material" and "sentimental" component into this valuation. Still consumer theory is silent about whether a cash gift could also have such a "sentimental" component, so these papers does not quite demonstrate that in-kind gifts are superior.

Taking the theory at face value though, that the gifts people give indeed are inefficient, why might they do it? Waldfogel (2009) blames the "stigma" of giving cash. Dubner and Levitt (2007), writing about the (private) inefficiency of gift cards—many of which are never spent—decry a "social taboo" against cash, which "crushes the economist's dream of such a beautifully efficient exchange."

Economic theorists, seeking to understand precisely what such a stigma or taboo could mean, offer models interpreting it as signaling. In these models, the choice to use money—or not to use it—conveys some private information. In Camerer (1988), a romantic suitor gives an inefficient in-kind gift as a signal of his long-term commitment and interest. In Prendergast and Stole (2001b), the giver, by choosing an in-kind gift, takes a gamble that the recipient will like it, and so signals her confidence that she knows the recipient's preferences well. In such models, because choosing an in-kind gift over money is costly—and differentially costly to different types of givers—it can serve as a credible signal.

Similar thought goes into another prime area for research into the use and non-use of money: compensation for costly production.[4] Here too, basic economic theory makes a strong normative claim: if you want someone to do something costly or difficult, the most effective way to motivate them is to pay them, and the most efficient way to pay them is with something fungible like money.[5]

Against the strong predictions of the standard theory, some research suggests that payment and incentives can backfire, actually discouraging the recipient from performing the intended behavior. Gneezy and Rustichini (2000a) demonstrate that a day-care center, when it started fining parents for picking up their kids late, actually saw an increases in tardy collections.

4. Gift-giving and compensation need not be treated as separate practices. Mauss (2002 (1950)) influentially argued that what we term "gifts" are generally not given for free, but rather form part of an extended exchange.

5. Moreover, a common line of reasoning holds that if something is efficient we expect people will do it, thus eliding this normative prescription into a positive description.

Gneezy and Rustichini (2000b) show, with a lab experiment, that low compensation can lead to worse performance than zero compensation (although high-enough compensation led to higher performance than either). What lies behind these effects? Gneezy and Rustichini argue that fining people gives them moral license for bad behavior, because they think of the behavior as now acceptable, while paying for an activity crowds out intrinsic motivation to perform it. Kamenica (2012), surveying these and other results on the downside of incentives, points out that most can be interpreted through the lens of signaling and inference: the agent learns useful information from how the principal chooses whether and how to pay them.[6]

Bénabou and Tirole (2003) build a general model of signaling via the choice of compensation. In the model, a principal with private information sets a compensation scheme for an agent, who makes an inference about the private information and then takes some action. For example, a high wage may indicate that the task is difficult, which may demotivate the agent if it teaches her that the marginal return to effort is lower than the anticipated. Among other channels, the paper highlights a "trust effect", in which the principal by offering low-powered incentives, signals that she trusts the agent to do the right thing. This boosts the agent's self-confidence, and motivates her further. Although my model is not a special case of theirs,[7] my Proposition 2.7 carries a similar theme of signaling trust.

Bénabou and Tirole (2006) model signaling of a different sort that may motivate people to forgo using money. Responding to Titmuss (1970), who argued that a system of blood

6. He highlights two other channels for how incentives can influence behavior: reference points for loss aversion, and psychological "choking" under pressure.

7. In my model, after the compensation scheme is set in period 0, both the price-chooser (the principal, in Bénabou and Tirole's (2003) nomenclature) and the other individual take actions in a game. In Bénabou and Tirole (2003), only the agent takes action. For this reason, my trust signaling result holds even without satisfying their condition "that the principal has information about the agent or the task that the agent does not". Instead, the principal has information about her own beliefs about the agent, which influence the principal's play in the game, and hence the agent's behavior is affected by learning this information.

donation yields more and better quality blood than a system of blood sale,[8] they show that it is theoretically possible that introducing compensation for charitable acts could decrease the performance of those acts, violating the law of supply. They also show that introducing price may change the composition of who performs those acts. In this model, agents have heterogeneous intrinsic preferences for contributing to the social good and for their own consumption; they also care about their reputation as socially-oriented (i.e. a preference for "esteem"). When the compensation for blood (say) is zero, only those who are motivated by altruism and by this esteem-seeking donate blood. But if the compensation is positive, then some who donate are simply motivated by money. Thus, the inference of an outsider, upon learning that someone has given blood, is more muddied; the potential do-gooder therefore gets less of a reputational boost from doing good.

Ellingsen and Johannesson (2011) also employ esteem-seeking preferences to explain why individuals will provide in-kind help but not monetary help to their friends, e.g. offering to help move but not to pay for a mover. Their goal is similar to my paper, to "rationalize the widespread use of seemingly inefficient non-monetary giving". Agents in their model care about three things: their own payoff (net their cost of helping), their friend's payoff (i.e. altruism), and their reputation as altruistic. In order to drive the model, Ellingsen and Johannesson assume first that in-kind help is less efficient than cash, but that it is relatively more efficient for altruists. As they explain, "although most people may find it onerous to help out with moving, it is usually less onerous to spend such time with a person one feels altruistic towards than with other persons." In this way, friends have an absolute and comparative advantage in in-kind gifts relative to non-friends, and a motivation to demonstrate this. The authors extend the model to explain why friends refrain from asking for monetary gifts (but do ask for in-kind help), by assuming that people differentially value esteem based on who provides it; asking for money reveals oneself to be greedy, and

8. See also Arrow (1972), Becker (2006), and Slonim et al. (2014) on the economics of blood donation or selling. It would be only a slight exaggeration to say that a whole branch of economics—market design—

less worthy of impressing. This second channels is also present in their work on reciprocal generosity that arises from differential esteem (Ellingsen and Johannesson, 2008).

Beyond these works on signaling, a handful of papers study reasons to restrict monetary exchange even when the choice of the means of exchange does not convey information.

A third route economists take is simply to assume that money can or cannot be used, and to study and compare the resulting outcomes. Within monetary theory, a sub-literature studies whether money is "essential", meaning that higher-welfare equilibria can be sustained with money than without it. In the model of Kiyotaki and Wright (1989), money is essential: introducing a fiat currency (i.e. a medium of exchange with no intrinsic value) unambiguously improves welfare. Lagos and Wright (2005) instead find that money is not essential, because there exist gift-exchange equilibria (i.e. dynamic equilibria) that give higher welfare. Duffy and Puzzello (2014) test a version of this model experimentally. Their version has a monetary equilibrium as well as a continuum of gift-exchange equilibria, where cooperation is sustained by grim trigger strategies, punishing shirking with the contagious breakdown of cooperation by everyone in the economy. Some of these gift-exchange equilibria Pareto-dominate the monetary equilibrium. However, the best of these gift-exchange equilibria was not a good predictor of behavior in the lab; instead, subjects picked sub-optimal equilibria. In contrast, when subjects were allowed to use money, their behavior was well predicted by the monetary equilibrium. Money comes out as essential in practice, even if not necessarily so in theory.

A small literature studies other models of gift- or favor-trading. Möbius (2001) is a highly-cited, never-published paper in this area. In it, each period agents stochastically get an opportunity to provide a favor or develop a need for one. This opportunity is private information, but the cost and benefit of giving and receiving a favor are common knowledge. Money is unavailable, and agents are selfish. Möbius (2001) studies a particularly simple

"chips" equilibrium, in which agents keep track of the past with a simple statistic: if they grant a favor, they get a "chip" up to some finite limit, and if they receive a favor, they give up a "chip" down to the limit of zero. Agents have an incentive to provide favors because, in equilibrium, this gives the right to request one later. Equilibrium welfare turns out to be increasing in the number of chips. However, it is never first-best.

While Möbius (2001) uses this model to study relaying favors within a network, others have expanded it in a one-on-one setting. Abdulkadiroğlu and Bagwell (2013) study a repeated game where it is efficient for agents to invest with each other, but they must trust that they will get paid back after the investment yields its returns. They study reciprocity, and contrast it to a chips mechanism (which they explore in Abdulkadiroğlu and Bagwell (2012) by developing a discrete version of the Möbius (2001) model). They find that there are more efficient equilibria than the simple reciprocity-based chips equilibrium, in which a honeymoon period of high trust (i.e. unconditional investment) in the first period is followed by favor trading or symmetric punishment. Hauser and Hopenhayn (2008) study alternative equilibria to the Möbius (2001) model. Employing the recursive methods of Abreu et al. (1990), they find that the Pareto Frontier of Public Perfect Equilibria is self-generating and so the equilibrium is renegotiation proof. Equilibria that stay on the frontier resemble chips equilibria, except that the rate of exchange varies over time (i.e. inflation) and with the agents' relative wealth. Kalla (2010) studies a similar favor-trading environment to these, but where agents' discount factors are private information. Focusing on one-chip equilibria to keep the model tractable, he characterizes sufficient conditions for the patient player to signal their patience, which then supports a favor-exchange equilibrium.

Studying repeated favor-exchange in an environment where costs and benefits are stochastic, Neilson (1999) shows that that no equilibrium can be first-best: because the value of the relationship is less than the net benefit from the highest favor, some favors are simply too

arose to take seriously (and treat as exogenous) the prohibition of organ sale (Roth, 2015).

costly to be worth performing.

Kranton (1996) builds a model that features both a market sector with money, and an informal sector with favor exchange. She shows that when more people use money, it becomes harder to find a trading partner in the informal sector. In this way, the two sectors interact, as substitutes.

Friendship and favors is a fertile topic, with deep theoretical questions and a variety of human behaviors to catalog and understand. This present work seeks to advance our understanding by characterizing one model, to better understand the motives behind friends helping each other without compensation.

CHAPTER 1

THE MARGINAL VALUE OF COMPENSATION IN A BILATERAL TRADING PARTNERSHIP

1.1 Overview of the model

This model depicts a friendship or partnership between two people who feel altruism towards one another. In the model, an opportunity arises where one can help the other by providing a service. The seller has private information about the cost of providing this service, and the buyer about the benefit of receiving it. We use the model to study whether, and under what conditions, altruism would make them prefer to transact for free vs. at a positive price.

The two players engage in a simple bargaining game. They make draws of a stochastic cost and benefit, and decide whether or not to trade. If they trade, the buyer pays the price a fixed and price p to the seller.

We use the model to study the conditions under which players wish to use or forgo compensation. In particular, when is a player's welfare maximized at $p = 0$, and when at some positive p?

We first pose this question with roles already fixed, i.e. with the players having common knowledge of who is the buyer and who the seller. We then randomize the roles, and ask when would someone prefer zero vs. positive price, if she did not know if she would be giving or receiving help. The idea is to consider what rules or norms a person would wish to set up when they begin a partnership with someone, knowing that they might help each other later. Is it better to establish a partnership where friends do not offer or accept money from one another, simply requesting and granting help on a voluntary basis? Or is it better to set a norm of compensation?

In short, the model reveals that compensation is better in most circumstances. Players prefer a positive price over a zero price whether they are buyer or seller, or if they are randomized into the two roles. The only exception is that a player would prefer a price of zero if she were sufficiently selfish and her partner were sufficiently altruistic. This exception applies for a buyer as well for someone randomized into the roles, though not for a seller.

These results mean that while the model does allow for circumstances would someone would choose not to use money, the only such circumstances are when a selfish person is the one making this choice, so that she does not need to pay for favors provided by an altruistic partner. The model provides no support to the idea that altruists might prefer to deal without payment.

Two features distinguish this model from previous work. First, the players have altruistic preferences. They care about their own material payoff (i.e. consumption) and that of their partner. With these preferences, even at a price of $p = 0$, some trade can still takes place, if the benefit is high enough and the cost low enough. This suggests that forgoing compensation, and relying solely on altruism, might be a viable alternative to paying for services provided. Second, we study players' preferred price when they do not yet know which side of the market they will sit on. This is meant to represent friendships and other relationships where the two participants are on a more equal footing than in market transactions, where the buyer and seller know who they are. In this model, each person may be able to offer help to the other.

The chapter proceeds as follows. Section 1.2 sets up the model. Section 1.3 establishes equilibrium behavior. Section 1.4 studies when the buyer and seller prefer compensation, and Section 1.5 studies this preference for compensation when players do not yet know their roles.

1.2 Model Setup

1.2.1 Environment

There are two players, a buyer b and a seller s. In this bilateral trading game, the seller gets the opportunity to provide a service to the buyer. If they trade, they do so at a pre-agreed price p in the interval $P \equiv [0, 1]$; this price is common knowledge, and they treat it as exogenous.

At the start of the game, each player $i \in \{b, s\}$ draws a ***valuation*** v_i independently from an atomless distribution F_i with support $[0, 1]$. These distributions are common knowledge. The buyer's valuation v_b is her benefit from receiving the service; the seller's valuation v_s is her cost of providing it.

Because $v_s \geq 0$ and $v_b \geq 0$, it is common knowledge that the service is costly for the seller and beneficial to the buyer.[1] To fix ideas, for some results we will specialize to the tractable case of having both valuations distributed $\sim \texttt{Unif}\,[0, 1]$ (unit uniform): $f_b(\cdot) = f_s(\cdot) = 1$ everywhere within that support.

Each player, after observing her valuation (but not her partner's), decides whether to agree to trade. If they both agree, then they trade.[2]

A ***strategy*** for a player i is a mapping from valuations $v_i \in [0, 1]$ to a probability of agreeing to trade. It turns out that the only interesting strategies of this game are ***cutoff strategies*** (see Lemma 1.4). A cutoff strategy is characterized by a cutoff valuation $\kappa_i \in \text{supp}(F_i) =$

1. This assumption could be relaxed to explore activities that both parties enjoy doing, by allowing negative cost $v_s < 0$.

2. This environment can be considered a special case of a double auction (Chatterjee and Samuelson, 1983), where the buyer's action space of bids is restricted to $\{p, -\infty\}$ and seller's action space of asks is restricted to $\{p, \infty\}$; only if they both choose p does trade take place, and naturally it takes place at price p. See Section 3.1 for a discussion of extensions of this paper's model to a more general mechanism.

$[0, 1]$: a buyer employing a cutoff strategy agrees to trade when $v_b \geq \kappa_b$, and likewise a seller agrees to trade when $v_s \leq \kappa_s$. That is, the buyer buys when her benefit is high, and the seller sells when her cost is low. When the two play cutoff strategies, the probability that trade takes place is

$$\mathbb{P}[v_s \leq \kappa_s] \cdot \mathbb{P}[v_b \geq \kappa_b] \quad = \quad F_s(\kappa_s) \cdot (1 - F_b(\kappa_b)),$$

which is $\kappa_s \cdot (1 - \kappa_b)$ in the case of unit-uniformly distributed valuations.

For most of the analysis, we treat these roles (buyer and seller) as fixed. But in Section 1.5, we consider a partnership where players are randomized into roles, and we ask what price p someone would prefer if she did not yet know her role. This perspective captures incentives to set the norms at the start of a partnership, when the two recognize they may find themselves giving or receiving help.

1.2.2 Payoffs

Each player cares about her own ***material payoff*** and her partner's. Material payoff refers to benefits or costs experienced directly (essentially, consumption). In this partnership, material payoff comes from the service provided and the payments made. That is, if the two players trade at price p with valuations v_b and v_s, then the buyer gets material payoff of $v_b - p$ and the seller gets material payoff $p - v_s$. If they do not trade, then both get zero material payoff.

A player i's ***altruism*** α_i governs how much she cares about her partner's material payoff relative to her own. If i experiences material payoff y_i and j experiences y_j, then i's utility is $y_i + \alpha_i y_j$. These are simple altruistic preferences, following Becker (1991). Their altruism parameters α_b and α_s are common knowledge in this game. (We relax this assumption in Chapter 2.) We assume that $\alpha_i \in [0, 1]$, meaning that each player (weakly) cares about the

other, though not more than about herself.

In this partnership, if they do not trade, they get zero utility, while if they do trade, their (ex-post) ***utilities*** are

$$u_b(v_b, v_s) = (\quad v_b - p) + \alpha_b(-v_s + p)$$
$$u_s(v_s, v_b) = (-v_s + p) + \alpha_s(\quad v_b - p).$$

Although players wish to maximize their utility, Players seek maximize their expected utility, given the information they have. Since they do not know their partner's valuation, they take expectations over it. When the players employ cutoff strategies κ_b and κ_s, their (interim) ***expected utilities*** are

$$U_b(v_b, \kappa_s) = \int_0^{\kappa_s} u_b(v_b, v_s)\,\mathrm{d}F_s(v_s) \tag{1.1}$$
$$U_s(v_s, \kappa_b) = \int_{\kappa_b}^1 u_s(v_s, v_b)\,\mathrm{d}F_b(v_b)\,. \tag{1.2}$$

At the outset of the partnership, the players do not know their own draw of valuation either. Therefore, given their cutoff strategies κ_b and κ_s, their ***ex ante expected utilities*** are

$$\mathcal{U}_b(\kappa_b, \kappa_s) = \int_{\kappa_b}^1 U_b(v_b, \kappa_s)\,\mathrm{d}F_b(v_b) = \int_{\kappa_b}^1 \int_0^{\kappa_s} u_b(v_b, v_s)\,\mathrm{d}F_s(v_s)\,\mathrm{d}F_b(v_b) \tag{1.3}$$
$$\mathcal{U}_s(\kappa_s, \kappa_b) = \int_0^{\kappa_s} U_s(v_s, \kappa_b)\,\mathrm{d}F_s(v_s) = \int_0^{\kappa_s} \int_{\kappa_b}^1 u_s(v_s, v_b)\,\mathrm{d}F_b(v_b)\,\mathrm{d}F_s(v_s)\,. \tag{1.4}$$

In the subsequent analysis, we study how these ex ante expected utilities depend on p when both players employ equilibrium strategies.

1.3 Equilibrium Behavior

1.3.1 Characterizing Equilibrium

We now derive the equations for best-response and equilibrium strategies.

In this game, each player chooses her action (agree to trade, or not) after she learns her realized valuation. She seeks to maximize her interim utility, best-responding to her belief about her partner's strategy; in equilibrium these beliefs are correct. Recall that a strategy is defined as mapping valuations to actions, so we can think of strategies as being chosen in first, before valuations are drawn.

Define a strategy profile as ***autarkic*** if trade occurs with zero probability, and as ***non-autarkic*** if trade occurs with positive probability. The probability of trade is $(1 - F_b(\kappa_b)) \cdot F_s(\kappa_s)$, so a pair of cutoff strategies is non-autarkic if $\kappa_b < 1$ and $\kappa_s > 0$, interior of the boundaries of the valuation support where $F_b(1) = 1$ and $F_s(0) = 0$. The best-response equations determine if a non-autarkic strategy profile is an equilibrium:

Proposition 1.1 (Non-Autarkic Equilibrium Cutoffs)**.**
A pair of cutoffs (κ_b^*, κ_s^*) *with* $\kappa_s^* > 0$ *and* $\kappa_b^* < 1$ *forms a non-autarkic equilibrium if it satisfies*

$$\kappa_b^* = (1 - \alpha_b)\, p + \alpha_b \, \frac{\int_0^{\kappa_s^*} v_s \, \mathrm{d}F_s(v_s)}{F_s(\kappa_s^*)} \tag{1.5}$$

$$\kappa_s^* = (1 - \alpha_s)\, p + \alpha_s \, \frac{\int_{\kappa_b^*}^1 v_b \, \mathrm{d}F_b(v_b)}{1 - F_b\Big(\kappa_b^*\Big)}. \tag{1.6}$$

The proof appears in Appendix A.1. To make sense of these equations, it helps to describe them in terms of conditional expectations.

Remark 1.2 (Best-Response Equations with Expectations Conditional on Trading)**.**
These best-response equations (1.5)–(1.6) *characterize the cutoffs that make each player indifferent to trading, meaning that they satisfy*

$$0 = U_b\big(\kappa_b^*, \kappa_s^*\big) = \mathbb{E}\Big[u_b\big(\kappa_b^*, v_s\big) \mid v_s \leq \kappa_s^*\Big] \cdot \mathbb{P}\big[v_s \leq \kappa_s^*\big]$$
$$0 = U_s\big(\kappa_s^*, \kappa_b^*\big) = \mathbb{E}\Big[u_s\big(\kappa_s^*, v_b\big) \mid v_b \geq \kappa_b^*\Big] \cdot \mathbb{P}\big[v_b \geq \kappa_b^*\big],$$

where we have expanded out interim utility, using that it is nonzero only if one's partner chooses to trade.

Due to linearity of preferences, the expectations operator $\mathbb{E}\big[\cdot \mid$ *partner agrees to trade*$\big]$ *and the utility functions commute, so the solutions work out to*

$$\kappa_b^* = (1 - \alpha_b)\, p + \alpha_b\, \mathbb{E}\big[v_s \mid v_s \leq \kappa_s^*\big] \tag{1.7}$$
$$\kappa_s^* = (1 - \alpha_s)\, p + \alpha_s\, \mathbb{E}\big[v_b \mid v_b \geq \kappa_b^*\big], \tag{1.8}$$

which is an alternative but equivalent formulation of (1.5)–(1.6)*.*[3]

Having characterized non-autarkic equilibrium, we turn to conditions for equilibrium existence and uniqueness.

Proposition 1.3 (Equilibrium Existence and Uniqueness)**.**
An autarkic equilibrium exists for any values of parameters $\alpha_b, \alpha_s, p \in [0, 1]$.

A non-autarkic equilibrium exists unless either

- $p = 0$ *and* $\alpha_s = 0$*, or*
- $p = 1$ *and* $\alpha_b = 0$.

3. Note that this second version is even well-defined at edge cases $\kappa_s^* = 0$ and $\kappa_b^* = 1$.

The non-autarkic equilibrium is unique if, in addition, distributions F_b and F_s, and altruism parameters α_b, α_s jointly satisfy

$$\alpha_b \cdot \alpha_s \cdot \frac{d}{dx}\mathbb{E}_{v_b \sim F_b}\big[v_b \mid v_b \geq x\big] \cdot \frac{d}{dy}\mathbb{E}_{v_s \sim F_s}\big[v_s \mid v_s \leq y\big] \leq 1 \tag{1.9}$$

for all $x \in [0,1], y \in [0,1]$.

The proof appears in Appendix A.1. It consists of three parts.

First, autarky is always an equilibrium because a player facing a partner who never trades is ensured zero utility, which means refusing to trade is indeed a best-response for her (just like any other strategy). Second, under the conditions given, the non-autarky equilibrium exists by a simple bounding argument (laid out in Lemma A.1): so long as the price p is interior to $(0,1)$, or players put at least some altruistic weight on each other's material utility, then their best response cutoff will be pulled toward the interior of their valuation support, which means that they trade with positive probability.

Third, for the uniqueness result, condition (1.9) ensures that the best-response mapping is a contraction. For this condition to hold, either the players' altruism must be weak, or their expectation of their partner's valuation conditional on trade must not change too fast with the cutoff.

1.3.2 Properties of Non-Autarkic Equilibrium

Before proceeding with studying how players' equilibrium utility depends on the price p, we make some remarks on their behavior in this equilibrium.

Cutoff Strategies

Earlier, I referred to cutoffs "the only interesting strategies of this game". The following result justifies this claim.

Lemma 1.4 (Cutoff Strategies)**.**
In any non-autarkic equilibrium, players employ cutoff strategies.

The proof, appearing in Appendix A.1, is simply a generalization of the proof of Proposition 1.1. For a player i, regardless of whether partner j plays a cutoff strategy or some other strategy, the best response is a cutoff strategy based on $\mathbb{E}\Big[v_j \mid j \text{ agrees to trade}\Big]$.

Strategic Complementarity

A higher cutoff strategy has different implications by a buyer compared to a seller. For a buyer, a higher cutoff means asking *fewer* favors, but higher-benefit ones. For a seller, it means granting *more* favor requests, including higher-cost ones.

Proposition 1.5 (Strategic Complementarity)**.**
Best-response cutoffs κ_b and κ_s are strategic complements, strictly so when altruism is strictly positive.

Proof of Proposition 1.5 (Strategic Complementarity).
The buyer's best-response κ_b^* (1.7) is increasing in the seller's cutoff κ_s^*, and the seller's best-response (1.8) is increasing in the buyer's cutoff κ_b^*. These are strictly increasing if, respectively, α_b and α_s are strictly positive. □

This strategic complementarity occurs because the players care about one another. If the seller believes that, when the buyer asks for help, her benefit v_b must be high (i.e. above a

high cutoff κ_b), then she would be willing to help even at high cost v_s to herself. Likewise, if the buyer believes that the seller would be willing to help even at high cost v_s (i.e. below a high cutoff κ_s), then she would be more selective and ask for help only when her benefit v_b is high.

Note that this strategic complementarity occurs even though preferences (ex-ante utility equations (1.3)–(1.4)) do not exhibit global increasing differences. If the cross-derivatives of ex-ante utility with respect to the two strategies were globally positive, it would immediately imply that their cutoffs are strategic complements.[4] However, they are only locally positive in the neighborhood of best-response strategies. [5]

Comparative Statics

Proposition 1.6 (Comparative Statics on p)**.**
Equilibrium cutoffs κ_b^* and κ_s^* are increasing in price p, strictly so if at least one player's altruism parameter is < 1.

The proof, which appears in Appendix A.1, comes from a straightforward examination of the first order conditions. It means that with anything less than full altruism, which renders individuals insensitive to transfers, supply curves still slope upwards and demand curves still

4. To see that preferences do not have global increasing difference, differentiate ex-ante utility (1.3)–(1.4) as

$$\frac{d^2}{d\kappa_b\, d\kappa_s}\mathcal{U}_b(\kappa_b,\kappa_s) = -u_b(\kappa_b,\kappa_s)\, f_b(\kappa_b)\, f_s(\kappa_s)$$
$$\frac{d^2}{d\kappa_s\, d\kappa_b}\mathcal{U}_s(\kappa_s,\kappa_b) = -u_s(\kappa_s,\kappa_b)\, f_s(\kappa_s)\, f_b(\kappa_b)\,,$$

and note that these may be positive or negative, depending on the values of κ_b and κ_s

5. To see that the cross-derivatives of ex-ante utility are locally positive around best-responses, e.g. for the buyer, compare $u_b(\kappa_b,\kappa_s)$ to $u_b\Big(\kappa_b, \mathbb{E}\big[v_s \mid v_s \le \kappa_s\big]\Big)$. The former, which appears in $\frac{d^2}{d\kappa_b\, d\kappa_s}\mathcal{U}_b(\kappa_b,\kappa_s)$, is b's ex-post utility evaluated at both cutoffs, while the latter is b's interim utility. When κ_s is a best-response to κ_b, the latter is zero by b's first order condition, so the former is negative because s's marginal valuation κ_s is always above her average valuation $\mathbb{E}\big[v_s \mid v_s \le \kappa_s^*\big]$ and because b's ex-post utility $u_b(v_b, v_s)$ is decreasing in s's valuation v_s.

slope downwards.

While comparative statics with respect to price are monotonic, we can only make a more limited statement about comparative statics with respect to altruism.

Proposition 1.7 (Comparative Statics on Altruism)**.**
At $p = 0$, equilibrium cutoffs κ_b^* and κ_s^* are increasing in both player's altruism parameters.

The proof appears in Appendix A.1. It relies on strategic complementarity and that the best-responses are increasing in altruism at $p = 0$.

Proposition 1.7 applies when $p = 0$, and it is also holds for p close to 0 (if altruism is high enough to make it hold strictly at $p = 0$). However, at sufficiently high p, the effect of altruism on cutoffs turns negative. Since a player i's best-response is $(1 - \alpha_i)p + \alpha_i \mathbb{E}\Big[v_j \mid j \text{ agrees to trade}\Big]$, a higher altruism re-weights i's incentives from the transfer p towards her partner j's valuation. If the price is higher than i's trade-contingent expectation of j's valuation, this change decreases her cutoff rather than increasing it.

1.3.3 Uniform Example

To push beyond what can be said in the general case will require specializing to a specific valuation distribution. We use the unit-uniform distribution, taking $v_b, v_s \overset{\texttt{iid}}{\sim} \texttt{Unif}\,[0,1]$. Under this distribution, within the support $[0,1]$, the PDFs are flat ($f_b(v_b) = f_s(v_s) = 1$) and the CDFs are the identity function ($F_b(x) = F_s(x) = x$). The probability of trade, given cutoffs κ_b and κ_s, is $\mathbb{P}[v_b \geq \kappa_b] \cdot \mathbb{P}[v_s \leq \kappa_s] = (1 - \kappa_b) \cdot \kappa_s$, and the players' valuations conditional on agreeing to trade are $\mathbb{E}\big[v_b \mid v_b \geq \kappa_b\big] = \frac{1+\kappa_b}{2}$ and $\mathbb{E}\big[v_s \mid v_s \leq \kappa_s\big] = \frac{0+\kappa_s}{2}$.

Because the distributions are linear in cutoff, ex ante utility (1.3)–(1.4) is quadratic in own

and partner cutoffs (and a third-order polynomial in cutoffs overall):

$$\mathcal{U}_b(\kappa_b, \kappa_s) = \left(\frac{1+\kappa_b}{2} - \alpha_b \frac{0+\kappa_s}{2} - (1-\alpha_b)\, p \right) \cdot \kappa_s \cdot (1-\kappa_b) \tag{1.10}$$

$$\mathcal{U}_s(\kappa_s, \kappa_b) = \left(-\frac{0+\kappa_s}{2} + \alpha_s \frac{1+\kappa_b}{2} + (1-\alpha_s)\, p \right) \cdot (1-\kappa_b) \cdot \kappa_s. \tag{1.11}$$

The best-response equations (1.5)–(1.6) are linear in cutoff as well:

$$\kappa_b^* = \alpha_b \frac{0+\kappa_s}{2} + (1-\alpha_b)\, p \tag{1.12}$$

$$\kappa_s^* = \alpha_s \frac{1+\kappa_b}{2} + (1-\alpha_s)\, p. \tag{1.13}$$

This system of linear equations is then easily solved for the unique non-autarkic equilibrium:

$$\kappa_b^* = \frac{\frac{\alpha_b\alpha_s}{4} + p\left(1 - \frac{\alpha_b}{2} - \frac{\alpha_b\alpha_s}{2}\right)}{1 - \frac{\alpha_b\alpha_s}{4}} \tag{1.14}$$

$$\kappa_s^* = \frac{\frac{\alpha_s}{2} + p\left(1 - \frac{\alpha_s}{2} - \frac{\alpha_s\alpha_b}{2}\right)}{1 - \frac{\alpha_s\alpha_b}{4}}. \tag{1.15}$$

These closed-form expressions make it simpler to investigate our main question of what price p players would prefer. While this distribution is a specific one, it serves as a useful baseline since it has no spikes in the probability of specific valuations within the distribution.

1.4 Compensation in the Partnership

With equilibrium behavior established, we now turn to our motivating question: under what circumstances would individuals in this model prefer to forgo money, and instead help each other for free?

Define the ***value of the partnership*** to a player b or s, given price p, to be her ex ante utility given (non-autarkic) equilibrium cutoffs:

$$W_b(p) = \mathcal{U}_b\big(\kappa_b^*(p)\,,\,\kappa_s^*(p)\,;p\big) \tag{1.16}$$

$$W_s(p) = \mathcal{U}_s\big(\kappa_s^*(p)\,,\,\kappa_b^*(p)\,;p\big)\,. \tag{1.17}$$

The term p appears as an explicit argument here to emphasize that utility and strategies depend on price. These expressions also depend implicitly on altruism parameters; when useful we will make this dependence explicit by writing the value of the partnership as $W_i\Big(p;\alpha_i,\alpha_j\Big)$ (for $i,j\in\{b,s\}$).

In order to understand when $W_i(p)$ is maximized at $p=0$ and when at a positive p, we examine its local behavior around $p=0$. Define the ***marginal value of compensation*** as $\frac{d}{dp}\Big|_{p=0} W_i\Big(p;\alpha_i,\alpha_j\Big) \equiv W_i'\Big(0;\alpha_i,\alpha_j\Big)$, for $i\in\{b,s\}$. If $W_i'(0)>0$, then i would prefer a price higher than p. If $W_i'(0)<0$, then a small increase in price would make i worse off, implying that $p=0$ is a local maximum (and possibly a global maximum as well).

When is the marginal value of compensation negative, and when is it positive? Using the model with generally distributed valuations, we can answer this question definitively for the seller and give sufficient conditions for it to be determinable for buyer.

Proposition 1.8 (Marginal Value of Compensation to Seller and Buyer)**.**
For the seller:

- $W_s'(0;\alpha_s,\alpha_b)\geq 0$ *for all parameter values* $\alpha_s,\alpha_b\in[0,1]$.

For the buyer:

- $W_b'(0;\alpha_b,\alpha_s)\geq 0$ *for* α_s *sufficiently close to* 0*, with strict inequality if* $f_s(0)>0$.

- $W_b'(0;\alpha_b,\alpha_s) < 0$ *for* $\alpha_s = 1$ *and* $\alpha_b = 0$*, if the valuation distributions satisfy the following condition:*

$$\frac{f_s\big(\mathbb{E}[v_b]\big)}{F_s\big(\mathbb{E}[v_b]\big)} \cdot \frac{f_b(0)}{1-F_b(0)} \cdot \big(\mathbb{E}_{v_b\sim F_b}[v_b]\big)^2 < 1. \tag{1.18}$$

The proof appears in Appendix A.1. These results rest on the interplay of two forces: an inframarginal effect from changing the transfer for trades that already take place, and a marginal effect from driving more or fewer trades.

For the seller, both effects push in the same direction, to increase utility. The inframarginal effect of a higher price is beneficial to the seller, since although she may have some altruism, she still prefers to receive more money; the marginal effect is to deter low-benefit buyers from trading (these are the lowest-value trades for the seller).

For the buyer, the two forces work in opposition, which is why the slope $W_b'(0)$ is ambiguous. When $\alpha_s = 0$, the seller's cutoff is simply $\kappa_s^* = p$, implying that literally no sellers would be willing to sell at $p = 0$. Because of this, the inframarginal effect to the buyer is zero, while the marginal effect is to induce some (low-cost) sellers to sell. So, when facing a selfish seller, the buyer unambiguously prefers a higher price over a zero price.

But the buyer's calculation changes when α_s is large (i.e. close to 1). With this much altruism, some low-cost sellers are willing to trade even with no compensation. This makes the inframarginal effect harmful to the buyer, since she has to pay a higher price in the positive-probability event that the seller sells. In contrast, the marginal effect helps the buyer by drawing in additional sellers. Inequality (1.18) is satisfied when the inframarginal effect wins out.

Proposition 1.8 leaves ambiguous the buyer's marginal value of compensation in intermediate cases. However, under uniformly distributed valuations, we can fully characterize its

sign.

Proposition 1.9 (Marginal Value of Compensation under Uniform Valuations)**.** *Suppose* $v_b, v_s \sim \texttt{Unif}\,[0,1]$*. For the seller:*

- $W_s'(p;\alpha_s,\alpha_b) > 0$*, unless* $\alpha_b = \alpha_s = 1$*, in which case* $W_s'(p;\alpha_s,\alpha_b) = 0$*.*

For the buyer, there exists an altruism threshold function $\zeta_b(\cdot)$ *such that*

- $W_b'(0;\alpha_b,\alpha_s) < 0$ *if* $\alpha_s > \frac{2}{3}$ *and* $\alpha_b < \zeta_b(\alpha_s)$*,*
- $W_b'(0;\alpha_b,\alpha_s) = 0$ *if* $\alpha_s \geq \frac{2}{3}$ *and* $\alpha_b = \zeta_b(\alpha_s)$*, and*
- $W_b'(0;\alpha_b,\alpha_s) > 0$ *if* $\alpha_s < \frac{2}{3}$*, or if* $\alpha_s \geq \frac{2}{3}$ *and* $\alpha_b > \zeta_b(\alpha_s)$*.*

This altruism threshold function $\zeta_b(\cdot)$ *is increasing in its argument, from* $\zeta_b\left(\frac{2}{3}\right) = 0$ *to* $\zeta_b(1) = 1$*.*

The proof appears in Appendix A.1, as does the buyer's threshold function $\zeta_b(\cdot)$ (A.18). The proof makes use of the closed-form equilibrium solution (1.14)–(1.15).

Figure 1 illustrates the region in (α_b, α_s)–space for which the buyer's marginal value of compensation $W_b'(p;\alpha_b,\alpha_s)$ is negative. The boundary between the positive and negative regions is the curve given by $\alpha_b = \zeta_b(\alpha_s)$.

Taken together, these results imply that if two friends have chosen to forgo money while knowing their buyer/seller role, this choice was not motivated by their altruism. Although higher altruism does reduce the weight of money in the players' preferences, it still matters to them (except in the extreme case where they are completely selfless, with $\alpha_b = \alpha_s = 1$). Even with some altruism, the seller still prefers to receive more, and the buyer would only forgo money (giving herself a discount) if she were sufficiently selfish and were taking advantage of a particularly altruistic seller.

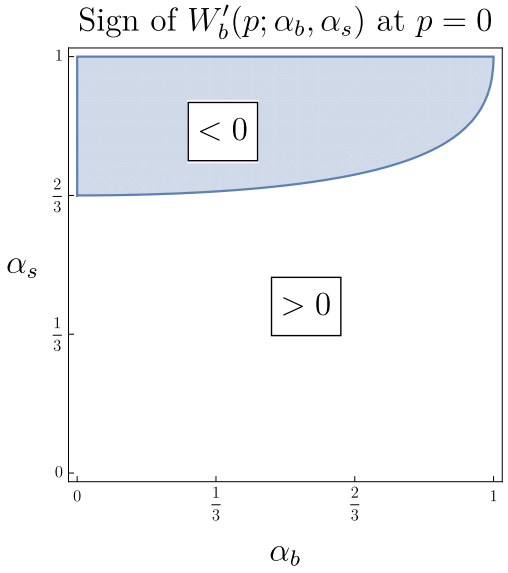

Figure 1: Altruism parameters in the blue region give rise to a negative marginal value of compensation to the buyer, i.e. $W_b'(0;\alpha_b,\alpha_s) < 0$. In this region of parameters, $p = 0$ is a local max to the buyer's value of the partnership. This region is characterized by a highly altruistic seller and a relatively selfish buyer, with $\alpha_b < \zeta_b(\alpha_s)$.

1.5 A Partnership with Random Roles

1.5.1 Motivation and Setup

In the previous section, we studied a partnership with fixed roles, where two individuals do not know their cost and benefit from a service, but do know who is providing it and who is enjoying it. However, many real-life relationships work differently, particularly those where people provide favors to each other without paying. Friends recognize that they might be called upon to help each other out when the need arises, but in advance they may not know who will be asking whom for help.

We therefore extend the model to describe a partnership with random roles, where the service provision could go in either direction. The model and notation need only slight modification. In this expanded model, there are still two individuals; we again use indices i and j for them. At the outset of their partnership, they set a price p that will be used in any future trades.

(For now, we do not specify a mechanism for choosing this price; we simply study their induced preferences over it.)

In this version of the model, the two players do not have fixed roles. Instead, they are randomly assigned the roles, with one becoming the buyer and the other the seller. Because of this, the role assignments are common knowledge. With these roles known, the two then play the previously-specified partnership game: they draw their valuations from distributions F_b and F_s, and then decide whether to trade.

We define the value of the partnership with random roles as the expectation across the two roles (b and s) of the fixed-role value of the partnership. That is, to a player i, the value is

$$W\Big(p;\alpha_i,\alpha_j\Big) = \frac{1}{2}W_b\Big(p;\alpha_i,\alpha_j\Big) + \frac{1}{2}W_s\Big(p;\alpha_i,\alpha_j\Big)\,. \tag{1.19}$$

This notation has no subscript on W, because the two players, devoid of roles, are distinguished only by their altruism parameters. In the right-hand side terms, $W_b\Big(p;\alpha_i,\alpha_j\Big)$ represents the value of the partnership to a buyer when the buyer's altruism is α_i and the seller's is α_j, and vice versa in $W_s\Big(p;\alpha_i,\alpha_j\Big)$.

1.5.2 The Marginal Value of Compensation with Random Roles

The random-role marginal value of compensation for a partnership is defined in the natural way, as $W'\Big(p;\alpha_i,\alpha_j\Big)$. It is simply the average of the marginal value of compensation under the two role assignments. If it is positive, it means a player would not choose to set a price of zero, because even a marginally higher level of compensation would give her more value in the partnership.

And indeed, the sign of the marginal value of compensation is described by similar rules to

the buyer's marginal value of compensation.

Proposition 1.10 (Marginal Value of Compensation with Random Roles)**.**
If $\alpha_j = 0$ *and* $\alpha_i < 1$*, then* $W'\Big(0; \alpha_i, \alpha_j\Big) > 0$*.*

If $\alpha_j = 1$ *and* $\alpha_i = 0$*, then* $W'\Big(0; \alpha_i, \alpha_j\Big) < 0$ *if the valuation distributions satisfy:*

$$\frac{f_s\big(\mathbb{E}[v_b]\big)}{F_s\big(\mathbb{E}[v_b]\big)} \cdot \frac{f_b(0)}{1 - F_b(0)} \cdot \big(\mathbb{E}_{v_b \sim F_b}[v_b]\big)^2 < 1. \tag{1.20}$$

The proof appears in Appendix A.1. Inequality (1.20) is actually the same condition that determined when $W_b'(0; 0, 1) < 0$ (inequality (1.18) from Proposition 1.8). This is because with $p = 0$ and with i so selfish, as a seller she is never willing to trade, so $W_s'(0; 0, 1) = 0$. This means that i's marginal value of compensation is driven entirely by her incentives when she is assigned the role of buyer.

Proposition 1.10 signs $W'(0)$ only in extreme cases of altruism parameters. To dig deeper, we again turn to the uniform distribution, and assume $v_b, v_s \sim \texttt{Unif}\,[0, 1]$.

Proposition 1.11 (Marginal Value of Compensation with Random Roles under Uniform Valuations)**.**
With random roles, there exists an altruism threshold function $\zeta(\cdot)$ *such that*

- $W'\Big(0; \alpha_i, \alpha_j\Big) < 0$ *if* $\alpha_j > \frac{2}{3}$ *and* $\alpha_i < \zeta\Big(\alpha_j\Big)$*,*
- $W'\Big(0; \alpha_i, \alpha_j\Big) = 0$ *if* $\alpha_j \geq \frac{2}{3}$ *and* $\alpha_i = \zeta\Big(\alpha_j\Big)$*, and*
- $W'\Big(0; \alpha_i, \alpha_j\Big) > 0$ *if* $\alpha_j < \frac{2}{3}$*, or if* $\alpha_j \geq \frac{2}{3}$ *and* $\alpha_i > \zeta\Big(\alpha_j\Big)$*.*

This altruism threshold function $\zeta(\cdot)$ *is increasing in its argument, from* $\zeta\Big(\frac{2}{3}\Big) = 0$ *to* $\zeta(1) = 1$*.*

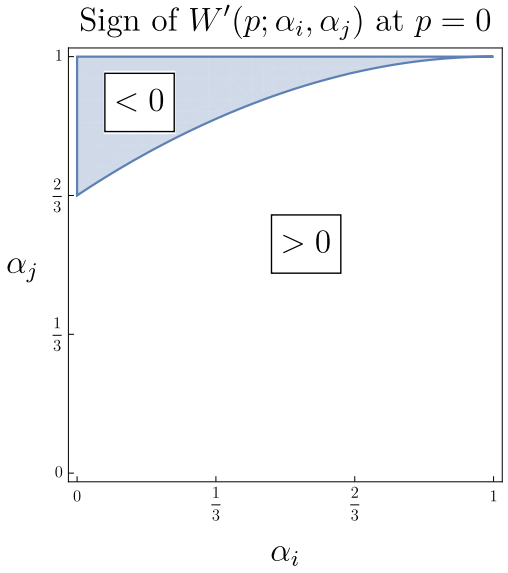

Figure 2: Altruism parameters in the blue region give rise to a negative marginal value of compensation with random roles, i.e. $W'\Big(0;\alpha_i,\alpha_j\Big) < 0$. In this region of parameters, $p=0$ is a local max to i's value of the partnership. This region is characterized by a highly altruistic j and a relatively selfish i, with $\alpha_i < \zeta\Big(\alpha_j\Big)$.

The proof appears in Appendix A.1 and the random-role threshold function $\zeta(\cdot)$ is defined in (A.19), while Figure 2 illustrates the region where $W'\Big(0;\alpha_i,\alpha_j\Big) < 0$. In this region, i prefers a zero price over a higher one. This random-role threshold function $\zeta\Big(\alpha_j\Big)$ is different than the buyer-specific one $\zeta_b(\alpha_b)$ from Proposition 1.9, as can be seen by comparing Figures 1 and 2. However, Propositions 1.9 and 1.11 are virtually identical, because at $\alpha_i = 0$the random-role marginal value of compensation is driven by the buyer's incentives, since a selfish seller i never sells.

A natural case of interest is a partnership of equals, where the two individuals feel the same altruism for one another. Equivalently, this could be interpreted as an externality equal to both, rather than just preferences.[6] In this situation, a positive price is preferable to a zero price.

Lemma 1.12 (Marginal Value of Compensation under Equal Altruism and Uniform Valu-

6. This equal-altruism assumption and externality interpretation is the approach taken by Bhaskar and

ations).

Suppose $v_b, v_s \sim \texttt{Unif}\,[0,1]$. *If* $\alpha_i = \alpha_j$, *then* $W'\Big(0; \alpha_i, \alpha_j\Big) \geq 0$, *with strict inequality except at* $\alpha_i = \alpha_j = 1$.

The proof appears in Appendix A.1. It shows that that if $\alpha_i = \alpha_j$ for $\alpha_j \in \left[\frac{2}{3}, 1\right]$, then $\zeta\Big(\alpha_j\Big) < \alpha_i$. By Proposition 1.11, this implies that α_i is in the "$W'(0) > 0$" region of Figure 2 , so the marginal value of compensation is nonnegative.

The implication of Lemma 1.12 is that in a truly equal partnership, where both parties have the same altruism, neither would prefer to forgo money. Each player is equally likely to be paying and receiving p, so the inframarginal effects cancel out between the two roles, and what remains is the marginal effect, that having a higher price than zero incentivizes more welfare-improving trades.

These results give us basic insight into how different altruism types differ in their preference for compensation.

Corollary 1.13 (Relative Sign of Marginal Value of Compensation).

Suppose $v_b, v_s \sim \texttt{Unif}\,[0,1]$. *If* α_L *and* α_H *are two values for* i*'s altruism that satisfy* $\alpha_L < \alpha_H$, *then:*

- *If* $W'\Big(0; \alpha_L, \alpha_j\Big) > 0$, *then* $W'\Big(0; \alpha_H, \alpha_j\Big) > 0$.
- *If* $W'\Big(0; \alpha_H, \alpha_j\Big) < 0$, *then* $W_i'\Big(0; \alpha_L, \alpha_j\Big) < 0$.

The proof appears in Appendix A.1. The logic is straightforward: Proposition 1.11's condition for $W'(0) \gtrless 0$ is an inequality $\alpha_i \gtrless \zeta\Big(\alpha_j\Big)$, and this condition is preserved if α_i is increased or decreased in the appropriate direction.

As with the fixed-role analysis from Section 1.4, these results indicate that we should not

Sadler (2017), who generalize the action space

expect more altruistic individuals to choose to forgo money. Rather, if anything, it is the more selfish types (low α_i) who would choose to partner up without using compensation. The model provides no support to the idea that altruists might prefer to deal without payment.

CHAPTER 2

UNCERTAINTY IN A BILATERAL TRADING PARTNERSHIP

2.1 Overview of Findings with Uncertainty

In Chapter 1, we modeled a friendship as a one-time bilateral trading game with altruism, to explore if altruists would prefer to forgo compensation, rather than set a positive price, for services they offer each other. We found one case where a player i would prefer to set a price of zero, namely when she is sufficiently selfish relative to her partner j. In this case, she chooses a price of zero because it lets her receive help without paying, outweighing a smaller loss from giving help without being paid for it.

We now explore a different reason why a player might choose to set $p = 0$: to signal private information. Even if, other things equal, she prefers a higher price, might she prefer a lower one in order to convince her partner of something about herself? We consider two possibilities of that "something": first that she is altruistic, and second that she trusts that her partner is altruistic.

In order to study such signaling, we need a model where there is initial uncertainty, so that some information is private and can be signaled via the choice of price. And because this price choice may not dispel all uncertainty, we must also introduce uncertainty into the trading game itself.

We therefore expand Chapter 1's model in two ways. First, we generalize the trading game to allow players to have uncertainty about their partner's altruism and beliefs. Second, we augment the model with an initial period where types are drawn from a common prior and then one player explicitly chooses a price.

This exploration yields two key results. Models such as Ellingsen and Johannesson (2011)

and Bénabou and Tirole (2006) suggest that an individual might forgo payment in order to signal their altruistic preferences. Our model does not support this story. The first result is that there does not exist a separating equilibrium in which the high-altruism type chooses a price of zero while the low-altruism type chooses some $\hat{p} > 0$. Although there can be benefit to convincing one's partner that one is altruistic, the selfish type gets more benefit from this reputation than the altruistic type, and so would mimic her; this separation therefore cannot be supported in equilibrium.

The second result concerns signaling trust. In a version of the model with correlated types—where different types hold different beliefs about their partner—there *can* exist a separating equilibrium where some types select $p = 0$, and in so doing signal that they trust that their partner is altruistic.

The separation works because altruism and trust can serve as a substitute for compensation, allowing people to offer mutual help. Someone who trusts her partner benefits from convincing her of this fact, so they can coordinate on a higher level mutual help. But someone who doubts her partner's altruism does not see much benefit in appearing to be trusting, since she knows that selfish people's actions are not very responsive to their beliefs about their partner. When this coordination benefit to the trusting type is high enough, it can outweigh the benefits from an efficient price seen in Chapter 1, and allow the more-trusting type to separate by choosing $p = 0$. However, to support it, this separating equilibrium requires quite extreme parameters, such as the more-trusting type assigning a probability of .995 that her partner is altruistic, while the less-trusting type assigns a probability of only .005 to this event.

Altogether, the model suggests that to the extent to which friendship is about mutually-beneficial transactions with one's friends, our forgoing of payment is not about signaling how much we care, but might be about signaling how much we trust that our friends care about us.

2.2 Setup of Model with Uncertainty

This model depicts two friends who help each other out, either for free or for a price. Formally, there are two individuals with altruistic preferences towards each other. The model is an extension of the one from Chapter 1, augmented to give players private information about altruistic preferences, and with an explicit pre-period of choosing the price p before they trade trading.

We refer to the two individuals as 1 and 2, and use indices i and j for them.[1]

They play a game with two periods. In the first period, 1 chooses and announces a price $p \in P \equiv [0, 1]$.[2] In the second period, the two play the bilateral trading game from Chapter 1, taking price p as exogenous. Throughout, each may have private information about her altruism parameter, which we refer to as her ***type***. They draw these types from a joint prior distribution π before the start of the game. After 1 chooses a price p, 2 makes an inference about 1's type, and they play the trading game with their updated beliefs. Thus, when 1 selects a price, she influences both a payoff-relevant parameter and 2's beliefs about her, both of which affect equilibrium play in the trading game.

Prior Distribution of Types

Before the game begins, the two players' altruism parameters $\alpha_1, \alpha_2 \in [0, 1]$ are drawn from a common prior distribution $\pi(\alpha_1, \alpha_2)$. For this model, we restrict attention to distributions with support of at most two types. When there are multiple types, the types of 1 are written as $\{L, H\}$ and the types of 2 are written as $\{\ell, h\}$. When the 1 types differ in altruism, we assume $\alpha_L \leq \alpha_H$, so that H consistently refers to the altruistic type and L the selfish

1. We use both notations $\{1, 2\}$ and $\{i, j\}$ because for most but not all parts of the model, the players are interchangeable, so the general notation helps avoid defining things twice.

2. In Section 2.6, we revise the rules so that randomly drawn player—either 1 or 2—chooses the price,

type. We use indices θ and φ to refer to types, with θ usually referring to player 1 and φ to 2.[3]

Importantly, the two players' types may be correlated. Thus, different types may not only have different altruism preferences, but may also hold different beliefs about their partner as well. A type θ of player i, after observing this draw of type, updates her belief about her partner j to

$$\pi\big(\varphi \mid \theta\big) = \frac{\pi(\theta, \varphi)}{\sum_{\varphi' \in \Phi_j} \pi\big(\theta, \varphi'\big)},$$

where we have used Φ_j to denote j's type space. This expression is the conditional probability $\mathbb{P}\big[j\text{'s type is } \varphi \mid i\text{'s type is } \theta\big]$, and is simply Bayes' rule.[4]

When we consider 2 having multiple types, we label them so h is more trusting that 1 is altruistic; that is, we assume $\pi\big(H \mid h\big) > \pi\big(H \mid \ell\big)$.

2.2.1 Choosing Price in Period 1

In period 1 of this game, player 1 unilaterally chooses a price p and announces it to 2. This price will be used in period 2 when they play a bilateral trading game.[5]

1 chooses this price to maximize her value of the trading game in period 2. This value

---

rather than just 1.

3. The exposition in this section will focus on the "2×2" version on the model, where both players have two types. It is straightforward to specialize this notation to the version with just one type, namely by setting its prior probability to 1 and the other types' prior probability to 0.

4. For example, if the prior distribution were

$$\begin{bmatrix} \pi(L,\ell) & \pi(L,h) \\ \pi(H,\ell) & \pi(H,h) \end{bmatrix} = \begin{bmatrix} .3 & .2 \\ .1 & .4 \end{bmatrix},$$

then if 1 learns that her type is H, she would update her belief about 2's type to be $\pi\big(h \mid H\big) = \frac{.4}{.1+.4} = .8$ and $\pi\big(\ell \mid H\big) = .2$.

5. As indicated in footnote 2, in Section 2.6 we modify the model to allow a randomly drawn player—

$W^\theta(p; \boldsymbol{\alpha}, \boldsymbol{\mu})$ is the payoff to type θ from the trading game where the types' altruism parameters are given by $\boldsymbol{\alpha} = \{\alpha_L, \alpha_H, \alpha_\ell, \alpha_h\}$ and their period-2 beliefs are given by $\boldsymbol{\mu} = \{\mu_L, \mu_H, \mu_\ell, \mu_h\}$, with $\mu_\theta(\varphi)$ referring to $\mathbb{P}\big[j\text{'s type is } \varphi \mid i\text{'s type is } \theta\big]$. 1's choice of p may affect her payoff $W^\theta(p; \boldsymbol{\alpha}, \boldsymbol{\mu})$ in two ways: directly via its first argument, and indirectly via changing 2's beliefs μ. We derive this value below in Section 2.2.2, but for now, treat it as an exogenous function that 1 tries to maximize by choosing p.

In period 1, player 1 may employ a pure or mixed strategy. If a type θ plays a pure strategy, choosing a single price, we write the strategy as p^θ. If she plays a mixed strategy, drawing from a probability distribution over price $p \in P$, we write it as $\sigma^\theta(\cdot)$, a function of price.

The equilibrium concept for period 1 is ***Perfect Bayesian Equilibrium***. A PBE consists of two objects that must obey two conditions. First, there is a strategy σ^θ by each type θ of 1, describing the probability distribution with which they choose price. Second, there is an updating rule $\pi\big(\theta \mid \varphi, p\big)$ by each type φ of 2, describing the posterior probability that φ assigns to player 1 being type θ after observing each price p. The two conditions these objects must mutually satisfy are:

Incentive Compatibility by 1: For both 1 types $\theta \in \Theta_1 = \{L, H\}$, for all $p \in P$,

$$\text{if } \sigma^\theta(p) > 0, \qquad \text{then } p \text{ maximizes } W^\theta\big(p; \boldsymbol{\alpha}, \boldsymbol{\pi}(p)\big)\,.^6 \tag{2.1}$$

Bayes Updating by 2: For both 2 types $\varphi \in \Phi_2 = \{\ell, h\}$, for all $p \in P$,

$$\text{if } \sum_{\theta' \in \Theta_1} \sigma^{\theta'}(p)\, \pi\Big(\theta' \mid \varphi\Big) > 0, \quad \text{then } \pi\big(\theta \mid \varphi, p\big) = \frac{\sigma^\theta(p)\, \pi\big(\theta \mid \varphi\big)}{\sum_{\theta' \in \Theta_1} \sigma^{\theta'}(p)\, \pi\big(\theta' \mid \varphi\big)}. \tag{2.2}$$

The IC condition (2.1) dictates that each type θ of 1 is maximizing her value, taking into account how her choice of p may change 2's beliefs about her. The Bayes Updating condition

either 1 or 2—to choose the price.

6. The bolded $\boldsymbol{\pi}(p)$ refers to the two players' profiles of posteriors, where 2 updates after 1 has publicly

(2.2) requires that each type of 2 update her belief following Bayes' rule where applicable (wherever the denominator is nonzero).

The Bayes Updating condition applies only at those p that 2 believes are played with positive probability (i.e. played with positive probability by some type θ' that 2 believes is drawn with positive probability).[7] It does not apply at every price p. Because it only constrains 2's beliefs after a p that is played with positive probability, PBE does not discipline beliefs after actions that are not played in equilibrium. Still, such beliefs matter for 1's IC condition as threats; when 1 considers deviating to a p not in the support of her strategy σ, she forecasts what 2 will think. PBE may admit choices and beliefs that would be ruled out by equilibrium refinements that specify what a "reasonable" out-of-equilibrium belief would need to be. However, this makes PBE a suitable equilibrium condition for negative results: if an outcome is ruled out as impossible under PBE, then under any sort of out-of-equilibrium beliefs, reasonable or unreasonable, it cannot occur.

While this game may admit many equilibria of the period-1 price choice, we study on those in which some types of 1 play $p = 0$ with positive probability, to understand friends' choices to forgo money.

2.2.2 Choosing to Trade in Period 2

The trading partnership game in period 2 progresses similarly to the one described in Section 1.2 of Chapter 1, with the modification that players' altruism preferences are no longer common knowledge. To account for multiple types, and to link to the period-1 p-choosing game, we lay out the model in full here in this section. (The only new notation introduced is μ for beliefs and the use of superscripts for types.)

---

chosen a price p but where 1 retains her prior since she has seen no new information. That is, $\boldsymbol{\pi}(p) := \left(\{\pi(\theta \mid \ell, p), \pi(\theta \mid h, p)\}_{\theta \in \Theta_1}, \{\pi(\varphi \mid L), \pi(\varphi \mid H)\}_{\varphi \in \Phi_2}\right)$.

7. In this game, unlike in standard treatments of signaling games such as Sobel (2009), the updating rule

In brief, there are two players 1 and 2, who each receive two pieces of information and then make one decision. First, the two players are publicly randomized into ***roles*** with equal probability, one becomes the buyer b, and the other the seller s. Each player i then draws their ***valuation*** v_i of the service; to the buyer this represents their benefit from receiving it, and to the seller, their cost of providing it. Their valuations are private information. The players then choose their action: whether or not to agree to trade. If they both agree, then trade takes place, with the seller providing the buyer a service, and the buyer paying the seller p.

For most of this section, we will refer to individuals by their role (b or s) and their type θ or φ, rather than which player they are (1 or 2). This game is symmetric with respect to their actual identity, so whether a particular player i is named 1 or 2 is not relevant at this point.

Types and Beliefs

Prior to playing this period-2 trading game, each player learns their type, which specifies their altruism and may convey information about their partner's type. We write $\mu\big(\cdot \mid \theta\big)$—or more compactly, $\mu_\theta(\cdot)$—for the belief held by a player i of type θ about her partner j's type. That is,

$$\mu\big(\varphi \mid \theta\big) \equiv \mu_\theta(\varphi) = \mathbb{P}\left[j\text{'s type is } \varphi \mid i\text{'s type is } \theta\right].$$

As the previous section outlined, these beliefs emerge as equilibrium objects in the period-1 game. They are formed by individuals updating from a common prior based on their own type, and then (for 2's belief) based on the choice of price p by 1 that she observes.

$\pi\big(\theta \mid \varphi, p\big)$ may depend on the type φ of 2. This situation is natural in this context, where different 2-types begin with different information about 1; indeed it is required by Bayes' rule at any p played by both types L and H. However, it also allows ℓ and h to differ in their beliefs following an out-of-equilibrium p.

However, it is important to calculate the value of this trading partnership game even at out-of-equilibrium beliefs, such as those that place probability zero on player 1 being a certain type, because the value of the partnership at these beliefs influences choices in period 1.[8]

We denote such a degenerate (or "pure") belief—one that puts full probability on a type θ—as $\mathbb{1}_\theta$. For example, if 2 holds belief $\mathbb{1}_H$, she is certain that 1's type is H. Such a belief will arise naturally after a separating equilibrium in period 1, if 2 observes 1 choosing a price that only H would choose in equilibrium.

Valuations and Strategies

As in Chapter 1, the valuations v_i are drawn from a commonly-known distribution F_i with bounded and connected support $[0, 1]$ for $i \in \{b, s\}$.[9] For tractability, for certain results we will specify that both valuations are distributed $\sim \texttt{Unif}\,[0, 1]$ (unit uniform), meaning the PDF is $f_b(\cdot) = f_s(\cdot) = 1$ everywhere within the support $[0, 1]$.

A ***strategy*** for a player i of type θ is a mapping from valuation $v_i \in [0, 1]$ to a probability of agreeing to trade. As in Chapter 1, we will focus on ***cutoff strategies***, as these are the only ones that will be played in equilibrium. A cutoff strategy is characterized by a cutoff valuation $\kappa_i^\theta \in \text{supp}(F_i)$. A buyer of type θ who plays cutoff κ_b^θ will agree to buy when $v_b \leq \kappa_b^\theta$, while a seller of type φ who plays cutoff κ_s^φ will agree to sell when $v_s \geq \kappa_s^\varphi$.

8. For this reason, we keep the notation μ for an arbitrary belief, and π for beliefs derived from updating through the game.

9. Note that here we use index i to refer to players by their role, since the model in period 2 treats the two players the same. The valuation and its distribution depend only on role, not on type. This model implicitly assumes that all types of the same individual/role draw the same valuation. Since different types of buyers, say, never interact with each other (only with the seller), there is no need to allow them to have different valuations. On a technical level, this assumption is useful because it allows expectation operators $\mathbb{E}_{v_i}$ and $\mathbb{E}_\theta$ to commute.

Payoffs

Each player has altruistic preferences toward her partner, caring about her own and her partner's ***material payoff***. A type θ's ***altruism*** $\alpha_\theta \in [0,1]$ is private information. If type θ of player i experiences material payoff y_i, and her partner j experiences material payoff y_j, then θ's utility is $y_i + \alpha_\theta y_j$.

In this game, if no trade takes place, then both players' material payoffs are zero, and since $0 + \alpha_\theta \cdot 0 = 0$, they get zero utility as well. If trade takes place at price p, with buyer and seller valuations v_b and v_s respectively, then their material payoffs are $v_b - p$ and $v_s + p$. Thus, if trade takes place between a buyer of type θ and a seller of type φ, their (ex-post) ***utilities*** are

$$\begin{aligned} u_b^\theta(v_b, v_s) &= (\quad v_b - p) + \alpha_\theta(-v_s + p) \\ u_s^\varphi(v_s, v_b) &= (-v_s + p) + \alpha_\varphi(\quad v_b - p). \end{aligned}$$

Players choose whether to trade in order to maximize their (interim) ***expected utility***, taking expectations of their partner's type and valuation. Suppose the buyer is type θ and the seller is type φ. Then, when both players employ cutoff strategies, their interim utilities are

$$U_b^\theta(v_b, \boldsymbol{\kappa}_s) = \mathbb{E}_{\varphi \sim \mu_\theta}\left[\int_0^{\kappa_s^\varphi} u_b^\theta(v_b, v_s)\,\mathrm{d}F_s(v_s)\right] \tag{2.3}$$

$$U_s^\varphi(v_s, \boldsymbol{\kappa}_b) = \mathbb{E}_{\theta \sim \mu_\varphi}\left[\int_{\kappa_b^\theta}^1 u_s^\varphi(v_s, v_b)\,\mathrm{d}F_b(v_b)\right], \tag{2.4}$$

where $\boldsymbol{\kappa}_s$ is the profile $\{\kappa_s^\varphi\}_{\varphi \in \Phi}$ of cutoffs for the seller types, and similarly $\boldsymbol{\kappa}_b$ is the profile $\{\kappa_b^\theta\}_{\theta \in \Theta}$ of the buyer types' cutoff.

Working backwards, before the buyer and seller learn their valuations, but after they learn their roles as buyer and seller, their ***ex ante expected utilities***, given their cutoffs, are

$$\begin{aligned}\mathcal{U}_b^\theta\left(\kappa_b^\theta, \boldsymbol{\kappa}_s\right) &= \int_{\kappa_b^\theta}^1 U_b^\theta(v_b, \boldsymbol{\kappa}_s)\mathrm{d}F_b(v_b) \quad = \int_{\kappa_b^\theta}^1 \mathbb{E}_{\varphi\sim\mu_\theta}\left[\int_0^{\kappa_s^\varphi} u_b^\theta(v_b, v_s)\mathrm{d}F_s(v_s)\right]\mathrm{d}F_b(v_b) \\ \mathcal{U}_s^\varphi\left(\kappa_s^\varphi, \boldsymbol{\kappa}_b\right) &= \int_0^{\kappa_s^\varphi} U_s^\varphi(v_s, \boldsymbol{\kappa}_b)\mathrm{d}F_s(v_s) = \int_0^{\kappa_s^\varphi} \mathbb{E}_{\theta\sim\mu_\varphi}\left[\int_{\kappa_b^\theta}^1 u_s^\varphi(v_s, v_b)\mathrm{d}F_b(v_b)\right]\mathrm{d}F_s(v_s). \quad (2.5)\end{aligned}$$

Although this notation has many superscripts, this ex ante expected utility is a straightforward extension of (1.3)–(1.4) from Chapter 1, expanded to simply take expectations over one's partner's type.

Given that all types play (non-autarkic) equilibrium cutoffs (as described in Section 2.3 below), the ***(fixed role) value of the partnership*** to the buyer or seller is simply expected ex ante utility:

$$\begin{aligned}W_b^\theta(p) &= \mathcal{U}_b^\theta\left(\kappa_b^{\theta*}(p)\,, \boldsymbol{\kappa}_s^*(p)\,; p\right) \\ W_s^\varphi(p) &= \mathcal{U}_s^\varphi\left(\kappa_s^{\varphi*}(p)\,, \boldsymbol{\kappa}_b^*(p)\,; p\right), \end{aligned} \quad (2.6)$$

where the starred cutoffs κ^* and cutoff profiles $\boldsymbol{\kappa}^*$ (in **bold**) are understood as equilibrium functions of p. These cutoffs κ, and hence these values W, also depend implicitly on the vectors of altruism $\boldsymbol{\alpha}$ and beliefs $\boldsymbol{\mu}$ of all types.

Lastly, working backwards to the beginning of period 2, define the ***(random role) value of the partnership*** as

$$W^\theta(p; \boldsymbol{\alpha}, \boldsymbol{\mu}) = \frac{1}{2}W_b^\theta(p; \boldsymbol{\alpha}, \boldsymbol{\mu}) + \frac{1}{2}W_s^\theta(p; \boldsymbol{\alpha}, \boldsymbol{\mu})\,.$$

This expression $W^{\theta}(\cdot)$ is the value of the partnership game in period 2, and it is the objective in period 1, when choosing the price.

Having derived the value of the partnership game in period 2, we have now completed the definition of the game with uncertainty. We now describe equilibrium of this game and study its implications.

2.3 Equilibrium Behavior

2.3.1 Equilibrium in the Partnership with Uncertainty

Some features of equilibrium are the same with this generalized game as for the trading game in Chapter 1. For one, in any equilibrium in which trade occurs, players employ cutoff strategies.

To extend the definitions from Section 1.3, we define a strategy profile as ***autarkic*** if trade occurs with zero probability (for all type realizations), and as ***non-autarkic*** if trade takes place with positive probability for at least some type realizations.

Equilibrium cutoffs are governed by a similar set of equations to (1.5)–(1.6):

Proposition 2.1 (Non-Autarkic Equilibrium Cutoffs)**.**
A profile of interior cutoffs $(\boldsymbol{\kappa}_b, \boldsymbol{\kappa}_s)$ *forms a non-autarkic equilibrium if they satisfy*

$$\kappa_b^{\theta} = (1-\alpha_{\theta})\, p + \alpha_{\theta} \cdot \frac{\mathbb{E}_{\varphi\sim\mu_{\theta}}\left[\int_0^{\kappa_s^{\varphi}} v_s \,\mathrm{d}F_s(v_s)\right]}{\mathbb{E}_{\varphi\sim\mu_{\theta}}\left[F_s\left(\kappa_s^{\varphi}\right)\right]} \tag{2.7}$$

$$\kappa_s^{\varphi} = (1-\alpha_{\varphi})\, p + \alpha_{\varphi} \cdot \frac{\mathbb{E}_{\theta\sim\mu_{\varphi}}\left[\int_{\kappa_b^{\theta}}^{1} v_b \,\mathrm{d}F_b(v_b)\right]}{\mathbb{E}_{\theta\sim\mu_{\varphi}}\left[1-F_b\left(\kappa_b^{\theta}\right)\right]} \tag{2.8}$$

for all buyer types θ and all seller types φ.

The proof appears in Appendix A.2. It follows the same reasoning as the proof of Proposition 1.1: preferences are linear, so the expected value of trading moves monotonically with one's own value. A seller will choose to trade when her cost is low enough to make this value nonnegative, and a buyer when her benefit is high enough.

Remark 2.2 (Best-Response Equations with Expectations Conditional on Trading)**.** *These equilibrium equations can be understood more clearly in terms of posterior beliefs about partner type, updating on the fact that one's partner being willing to trade. Equilibrium equations* (2.7)–(2.8) *can be written as*

$$\kappa_b^\theta = (1-\alpha_\theta)\,p + \alpha_\theta \cdot \mathbb{E}_{\varphi\sim\mu_\theta}\Big[\mathbb{E}\Big[v_s \mid v_s \le \kappa_s^\theta\Big] \mid s \text{ agrees to trade}\Big] \tag{2.9}$$

$$\kappa_s^\varphi = (1-\alpha_\varphi)\,p + \alpha_\varphi \cdot \mathbb{E}_{\theta\sim\mu_\varphi}\Big[\mathbb{E}\Big[v_b \mid v_b \ge \kappa_b^\varphi\Big] \mid b \text{ agrees to trade}\Big]. \tag{2.10}$$

In these expressions, the inner $\mathbb{E}[\cdot]$ takes expectations over the partner's valuation, and the outer over her expected type, both conditional on her agreeing to trade. As an example of this outer type updating, if H trades with high probability and L with low probability, then φ should update her belief from $\mu\big(H \mid \varphi\big)$ to a higher posterior $\mu\big(H \mid \varphi, 1 \text{ agrees to trade}\big)$.[10]

When a player draws a valuation exactly equal to the above cutoff (2.9) *or* (2.10)*, it makes her indifferent to trading. She takes into account the fact that, if her decision to trade matters at all, then her partner must have agreed to trade as well; she takes expectations accordingly. In other words, the best response is the cutoff for which, if i's valuation equals the cutoff,*

10. The precise posteriors take the following form: for a buyer of type θ and seller of type φ, these Bayesian posteriors (conditional on one's partner agreeing to trade) are

$$\mu\big(\varphi \mid \theta, \text{trade}\big) = \frac{\mu\big(\varphi \mid \theta\big)\cdot\big(F_s(\kappa_s^\varphi)\big)}{\sum_{\varphi'\in\Phi}\mu\big(\varphi' \mid \theta\big)\cdot F_s\Big(\kappa_s^{\varphi'}\Big)} \qquad \mu\big(\theta \mid \varphi, \text{trade}\big) = \frac{\mu\big(\theta \mid \varphi\big)\cdot\Big(1-F_b\big(\kappa_b^\theta\big)\Big)}{\sum_{\theta'\in\Theta}\mu\big(\theta' \mid \varphi\big)\cdot\Big(1-F_b\big(\kappa_b^{\theta'}\big)\Big)}.$$

her expected interim expected utility is zero, conditional on j wishing to trade.

Equations (2.9)–(2.10) *show the best response is a cutoff that is an altruism-weighted average of the price and one's partner's expected valuation, conditional on the partner wishing to trade. This is the same message as equations* (1.7)–(1.8) *from Chapter 1 without type uncertainty, only now with a more nuanced understanding of that expectation and conditioning.*

Equilibrium existence remains straightforward in this model with uncertainty:

Proposition 2.3 (Equilibrium Existence with Uncertainty)**.**
An autarkic equilibrium exists for any parameter values.

A non-autarkic equilibrium exists if $0 < p < 1$, or if $\alpha_\theta > 0$ for all types.

The proof appears in Appendix A.2. This result is analogous to the existence part of Proposition 1.3, and its proof runs along the same lines.

Although equilibrium existence is easy to prove, compared to Chapter 1, here with multiple types, it is more difficult to prove the uniqueness of non-autarkic equilibrium. While it appears true that there are never multiple non-autarkic equilibria to the trading game—all parameters tested return a unique non-spurious solution to the equilibrium equations—it remains unproven. The issue is that the best-response mapping is not a contraction—even in the simple case where valuations are distributed unit-uniformly—so standard tools like the Contraction Mapping Theorem cannot be applied.[11] Still, it may be possible to prove uniqueness using other approaches.

11. For example, with unit-uniformly distributed valuations and with two types $\{L, H\}$ of seller, it can be shown that at $\kappa_s^L = 0$ buyer's best response cutoff is $\left.\frac{d\kappa_b}{d\kappa_s^L}\right|_{\kappa_s^L=0} = -\frac{\alpha_b}{2} \cdot \frac{\mu(L)}{\mu(H)}$. This expression is negative and can have arbitrarily high magnitude, if b's belief μ puts sufficient weight on L relative to H. One the other hand, it *can* be shown the each player θ has best-response cutoff with slope of less than their altruism α_θ, which is < 1. However, the Contraction Mapping Theorem (Stokey and Lucas, 1989) requires not just that the best-response mapping have a slope of < 1 (the "discounting condition"), but that it also have slope of ≥ 0 (the "monotonicity" condition).

2.3.2 Solving for Equilibrium

We solve for a non-autarkic equilibrium by finding a solution to the best-response equations in which all cutoffs are in $[0, 1]$, assuming that valuations are unit-uniformly distributed. In general, we solve it with a computer algebra system. The system of equations to be solved is the unit-uniform version of (2.7)–(2.8), for all types θ and φ:

$$\kappa_b^\theta = (1 - \alpha_\theta) \cdot p + \alpha_\theta \cdot \frac{\sum_\varphi \mu_\theta(\varphi)\, \kappa_s^\theta \cdot \frac{0+\kappa_s^\theta}{2}}{\sum_\varphi \mu_\theta(\varphi)\, \kappa_s^\theta} \tag{2.11}$$

$$\kappa_s^\varphi = (1 - \alpha_\varphi) \cdot p + \alpha_\varphi \cdot \frac{\sum_\theta \mu_\varphi(\theta)\, (1 - \kappa_b^\varphi) \cdot \frac{1+\kappa_b^\varphi}{2}}{\sum_\theta \mu_\varphi(\theta)\, (1 - \kappa_b^\varphi)}. \tag{2.12}$$

After cross-multiplying the denominators, this is a system of third-order polynomials. In simple cases, it can be solved in closed form by hand, such as the 2×1 environment (i.e. two types of one player, one of the other) with degenerate beliefs,[12] or even in the 2×2 environment when altruism parameters are set to 0 or 1.[13]

For this paper, the equilibrium equations in more complex cases were solved to give exact solutions using *Mathematica* 11.3. With p left as a variable, the 2×2 system of equations tends to be solved on the order of minutes if the several of the parameters are specified (e.g. particular values pre-filled in for the most of the terms α_θ and $\mu(\varphi)$), while the solver does not halt in less than a day if all parameters were left arbitrary (in the full 2×2 model).

12. Suppose that 2 has belief $\mu_2(H) = 1$ about 1's type. (This would occur if, in period 1, 2 inferred 1's type after observing a p that only H would choose.) In this case, to solve we apply 1×1 equilibrium solution (1.14)–(1.15) to cutoffs of H and 2, who are certain that they're facing one another. Then, we apply best response equations (1.12)–(1.13) to calculate how L responds to 2. The resulting equilibrium is

$$\kappa_b^H = -\frac{\alpha_H\alpha_2 - 2p\alpha_H\alpha_2 - 2p\alpha_H + 4p}{\alpha_H\alpha_2 - 4} \qquad \kappa_s^2 = \frac{2\,(-\alpha_2 + p\alpha_H\alpha_2 + p\alpha_2 - 2p)}{\alpha_H\alpha_2 - 4}$$
$$\kappa_b^L = -\frac{\alpha_L\alpha_2 - p(\alpha_L + \alpha_H)\alpha_2 - 2p\alpha_L + 4p}{\alpha_H\alpha_2 - 4}.$$

13. For special cases of the 2×2 equilibrium solution, see footnotes 21 and 22 in Section 2.6.2, and the proof of Proposition 2.7 in Appendix A.2.

When the solver does return an answer, it is an exact, closed-form solution, not a numerical approximation, though it is usually extremely long.

2.4 Price Choice with No Type Uncertainty (1×1)

With the model now established, we now explore its implications, to determine when someone would choose a price of zero in period 1. We start with the simplest case, the 1×1 game, with just one type of each player. In this setup, each player's altruism parameter is common knowledge. To be concrete, we further specialize to the case where valuations are distributed Unit-Uniform, i.e. $v_b, v_s \sim \texttt{Unif}\,[0,1]$.

The result here follows straightforwardly from the results of Chapter 1. There, we derived equilibrium behavior and values with common-knowledge types for what we now call the period-2 trading game: equations (1.14)–(1.15) give equilibrium cutoffs (once roles have been assigned), equation (A.20) characterizes the value of the patnership $W\Big(p; \alpha_i, \alpha_j\Big)$ in closed-form, and Proposition 1.11 answers when the value of the partnership $W\Big(p; \alpha_i, \alpha_j\Big)$ achieves a max (within $P = [0,1]$) at $p = 0$. Leveraging all this, solving the equilibrium in period 1, in which 1 chooses a price p, is simple. In short, she chooses the price that maximizes the value of the partnership. Therefore, the situations when she would choose $p = 0$ closely follow the results of Proposition 1.11:

Proposition 2.4 (1×1 Equilibrium p Choice)**.**
Let the valuations be distributed $\sim \texttt{Unif}\,[0,1]$. *In the* 1×1 *game, there exists an altruism threshold function* $\zeta(\cdot)$ *such that it is a PBE for* 1 *to choose* $p_1 = 0$ *if* $\alpha_2 \geq \frac{2}{3}$ *and* $\alpha_1 < \zeta(\alpha_2)$.

This altruism threshold function $\zeta(\cdot)$ *is increasing in its argument, from* $\zeta\Big(\frac{2}{3}\Big) = 0$ *to* $\zeta(1) = 1$.

The proof appears in Appendix A.2.

This threshold function $\zeta(\cdot)$, given in equation (A.22), is the same formula as (A.19), given in the proof of Proposition 1.11. Proposition 2.4 works because $p_1 = 0$ is an equilibrium choice precisely when it is a max of 1's value of the trading game. This value of the partnership $W(p; \alpha_1, \alpha_2)$ is quadratic in p and symmetric about $p = \frac{1}{2}$, so when it has 0 as a local max in $P = [0, 1]$, it has zero as a global max in P as well. Note, however, that $p = 1$ is equally well a max, and so choosing that price would be an equilibrium just the same.

Figure 2 illustrates the region of altruism parameters where $W'(0; \alpha_1, \alpha_2) < 0$, i.e. for which $p = 0$ is a max of $W(p; \alpha_1, \alpha_2)$; the boundary of the region is the curve $\alpha_1 = \zeta(\alpha_2)$.

Proposition 2.4 tells us that while there are some situations where 1 would choose $p_1 = 0$, these situations involve 1 being much more selfish than her partner 2. As in Chapter 1 the logic is that in these situations, 1 recognizes that since she is selfish, for most trades that occur, she will be playing the role of the buyer, and so she chooses a price of zero to enjoy a free services from 2 in these cases.

2.5 Price Choice with First-Order Uncertainty (2×1)

We now turn to the simplest case of incomplete information: where one player's altruism is private information, and the other's is common knowledge. In this setup, 1 has two possible types—L for low altruism and H for high altruism, with $\alpha_L < \alpha_H$—and 2 has just one type.[14]

Now that we have an environment with private information that might be signaled, we use it to test our question of interest: can the forgoing of compensation serve as a signal of what sort of partnership the two individuals are engaged in? To make this question theoretically concrete, we ask if there exists an equilibrium under which someone chooses not to use compensation, and in so doing, signals that they have high altruism.

14. This is known as "first-order" uncertainty, because while 2 is uncertain of 1's altruism, no one has

This question is in line with explanations like that of Bénabou and Tirole (2006) and Ellingsen and Johannesson (2011), where an agent sacrifices receiving compensation in order to prove their type (as altruistic or pro-social). But unlike those papers, here our answer is *no* – our model does not admit such a signaling equilibrium.

Instead, we find a result like that in the environment without uncertainty (i.e. the 1×1 game). In that setup, Proposition 2.4 showed that lower altruism of 1 is associated with choosing $p = 0$ in period 1, under equilibrium. Here, with uncertainty about 1's type, a similar lesson emerges.

Proposition 2.5 (2×1 – No separating equilibrium where only H chooses $p = 0$)**.**
Let the valuations be distributed $\sim \texttt{Unif}\,[0, 1]$. *Suppose there are two types of* 1 *with* $\alpha_L < \alpha_H$ *and one type of* 2.

Then, there does not exist a Perfect Bayesian Equilibrium where only H *chooses a price of zero, and where only* L *chooses some price* $\hat{p} > 0$. *That is, there does not exist a PBE where* $\sigma^H(0) > 0$ *and* $\sigma^L(0) = 0$, *and where for some* $\hat{p} > 0$, $\sigma^L(\hat{p}) > 0$ *and* $\sigma^H(\hat{p}) = 0$.

The proof appears in Appendix A.2.

Proposition 2.5 answers the question of whether choosing price zero could signal high altruism; the answer is that it cannot. The reason is that, roughly speaking, if parameters are such that the high-altruism type H gains from choosing a price of zero and getting a high reputation, then the low-altruism type L has even more to gain from such a reputation. If such a strategy profile were to constitute an equilibrium, incentive compatibility would demand that L prefer not to switch to price zero and high reputation $\mathbb{1}_H$, from some $\hat{p}$ carrying low reputation $\mathbb{1}_L$. This switch would lead to lower material utility for 2, since fewer welfare-enhancing trades would take place. But L has less altruism than H, and so cares less about 2's material utility, meaning that the gain from switching to $p = 0$ from

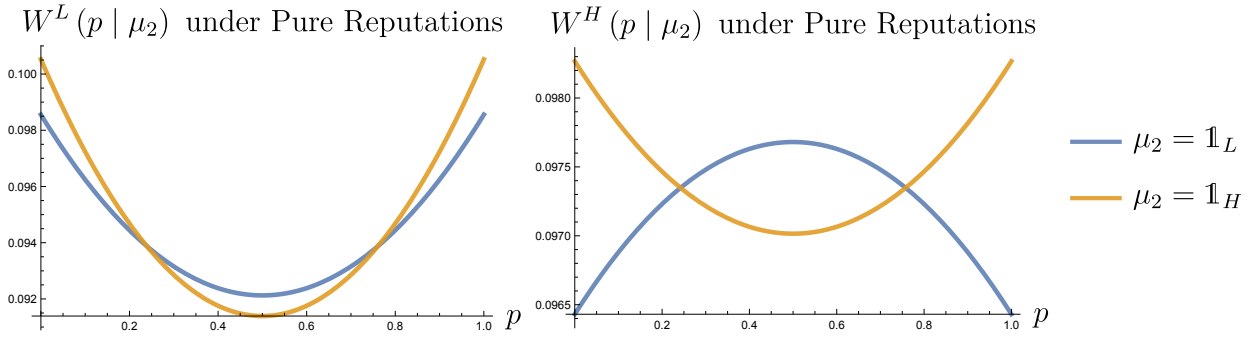

Figure 3: The value of the partnership to each type of 1, given parameter values $\alpha_L = .1, \alpha_H = .2, \alpha_2 = .8$, under low and high (pure) reputations $\mathbb{1}_L$ and $\mathbb{1}_H$. The curves show that an equilibrium cannot be supported where only H chooses price zero: if so, then Bayes rule for 2 would imply that $\pi\big(H \mid p = 0\big) = 1$ (2 infers the high type after seeing $p = 0$). But then L prefers to set $p = 0$ as well, since $W^L\big(0 \mid \mathbb{1}_H\big) > W^L\big(\hat{p} \mid \mathbb{1}_L\big)$ for all $\hat{p} > 0$.

some higher price would be *greater* for L than for H. Thus, if H does not want to choose some higher $\hat{p}$, L wouldn't want to either, and so this supposed equilibrium would actually break someone's Incentive Compatibility condition.

Figure 3 illustrates values $W^\theta(p)$ in one situation where incentive compatibility fails for L. It uses parameter values $\alpha_L = .1, \alpha_H = .2, \alpha_2 = .8$. In this case, H does prefer $p = 0$ with high reputation $\mathbb{1}_H$ over $p = \frac{1}{2}$ with low reputation. That is, $W^H\big(0 \mid \mathbb{1}_H\big) > W^H\Big(\frac{1}{2} \mid \mathbb{1}_L\Big)$. However, separation is not incentive-compatible, because if choosing $p = 0$ garners reputation $\mathbb{1}_H$, then L wants to choose it too: $W^L\big(0 \mid \mathbb{1}_H\big) > W^L\big(\hat{p} \mid \mathbb{1}_L\big)$ for all $\hat{p}$. Proposition 2.5 proves that such a separating equilibrium can never occur, and these parameters illustrate one example of that.

A consequence of Proposition 2.5 is that if equilibrium does involve separation at $p = 0$, it cannot involve the high type forgoing compensation with the low type insisting on it. Rather, it would have to be the low type forgoing money (for reasons articulated in Section 2.4 – to give herself a discount). The following strategy profile gives an example of this separation:

---

higher-order uncertainty about 2's belief. We explore second-order uncertainty in Section 2.6.

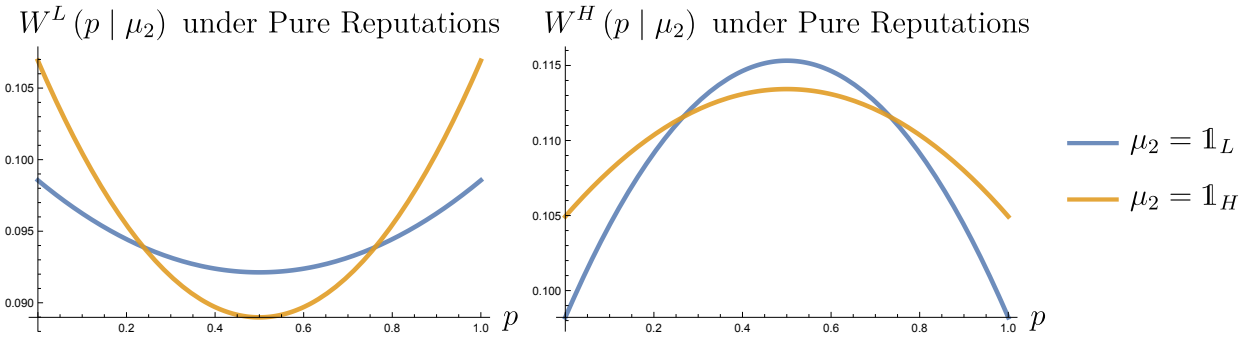

Figure 4: The value of the partnership to each type of of 1, given parameter values $\alpha_L = .1, \alpha_H = .5, \alpha_2 = .8$ from Example 2.6, under low and high (pure) reputations $\mathbb{1}_L$ and $\mathbb{1}_H$. The curves show that there exists an equilibrium with $p^L = 0$ and $p^H = .5$. While L prefers a price of 0 even at her true reputation $\mathbb{1}_L$, H is sufficiently altruistic that she prefers to set a price of .5 instead.

Example 2.6 (2×1 Separating Equilibrium with $p^L = 0$)**.**

With parameter values

$$\begin{aligned} \alpha_L &= .1 & \pi(L) &= .5 \\ \alpha_H &= .5 & \pi(H) &= .5, \\ \alpha_2 &= .8 \end{aligned}$$

there exists a separating 2×1 *equilibrium in which the* 1*-types' price choices and* 2*'s updating rules are*

$$\begin{aligned} p^L &= 0 & \pi\big(H \mid p \neq .5\big) &= 0 \\ p^H &= .5 & \pi\big(H \mid p = .5\big) &= 1. \end{aligned}$$

Figure 4 illustrates this situation. To see that incentive compatibility does hold (at one pair of actions, 0 and $\frac{1}{2}$), check that $W^H\big(0 \mid \mathbb{1}_L\big) < W^H\Big(\frac{1}{2} \mid \mathbb{1}_H\Big)$ and $W^L\big(0 \mid \mathbb{1}_L\big) > W^L\Big(\frac{1}{2} \mid \mathbb{1}_H\Big)$. As in Chapter 1's 1×1 example, here L prefers $p = 0$ because it enables her to pay less when she is the buyer, but H, with a higher altruism, does not share this predilection.

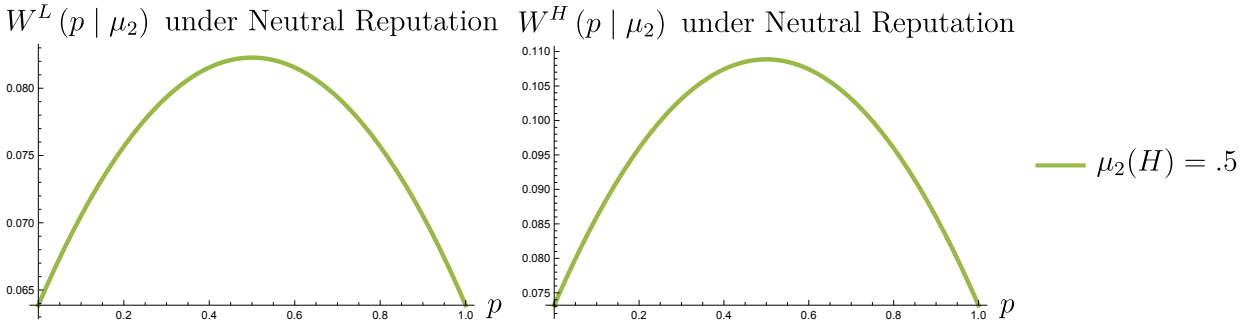

Figure 5: The value of the partnership to each type of of 1, given parameter values $\alpha_L = .1, \alpha_H = .5, \alpha_2 = .5$, under neutral reputation $\pi(L) = \pi(H) = \frac{1}{2}$. Both peak at $p = \frac{1}{2}$, and since in equilibrium 2's keeps her belief no matter what she observes, it is Incentive Compatibility for both types of 1 to simply play $p = \frac{1}{2}$ to maximize their value from the game.

At some parameter values, there also exist pooling equilibria, in which both types choose $p = \frac{1}{2}$ (see Figure 5) or both choose $p = 0$. However, thanks to Proposition 2.5, we know that if they separate, it cannot be with $\sigma^H(0) > 0 = \sigma^L(0)$. While a 2×1 setup may admit several different sorts of equilibrium, none involve the altruistic type separating by choosing a price of zero.

2.6 Price Choice with Second-Order Uncertainty (2×2)

In Sections 2.4 and 2.5, we explored the idea that friendship is about caring. We sought to understand why someone might choose to forgo using compensation with her friend, and with Propositions 2.4 and 2.5, we established that being more altruistic is *not* a viable explanation, nor is wanting to signal one's altruism.

However, friendship is about more than being altruistic. In this section, we explore further potential signaling motives that might lead someone to forgo compensation. In particular, we consider another important aspect to friendship: trust, where friends know they can count on each other to act in their interests. We now embellish the model slightly to determine if trust can explain friends forgoing compensation.

To address this question, we incorporate higher-order uncertainty into the environment. The individuals now differ not just in altruism as in Section 2.5, but also in their beliefs about each other. We call a type ***more trusting*** if she has a higher belief that her partner is the high type, a term we will use to refer to either the more altruistic or (recursively) more trusting type.[15]

Higher-order uncertainty can be complicated – players have not just beliefs about their partner's preferences (first-order beliefs), but also beliefs about their partner's beliefs about their own preferences (second-order beliefs), and so on. Fortunately, high-order uncertainty need not complicate the analysis too much: using a type-space approach (introduced by Harsanyi (1967, 1968a,b)), we can analyze such situations while only keeping track of a few additional variables. We simply continue referring to agents by their type θ (e.g. L or H), but now treat type as more general than simply a label for their altruism parameter. The only parameter to track is a player's belief about her partner's type; beliefs about preferences or beliefs—of however high an order—can be calculated as needed.

The simplest environment with interesting higher-order beliefs involves two types of each player, where the common-prior type distribution π involves correlation between types. In such an environment, we study the question: do there exist equilibria in which the more optimistic type chooses to forgo compensation (i.e. selects $p = 0$) in period 1, due to a signaling motive?

The answer turns out to be *yes*: a trusting player j benefits from persuading her partner i of this trust, while a less-trusting j does not. These differential benefits from having a trusting reputation drive the possibility of a separating equilibrium, in which the trusting type chooses to forgo compensation to send a costly signal of her trust.

The logic behind these motives as follows. If j believes that her partner i is altruistic, she

15. We call this "trust" because the more j cares about i, the more in line with i's preferences her actions

finds it useful to persuade i that she is trusting. This is because strategies are strategic complements when players are altruistic: e.g. if i believes j will request a favor only when the benefit is high, then an altruistic i will be more willing to provide favors even at high cost. However, a selfish i's strategy is not very responsive to what she believes j will do; rather, in determining her strategy she puts most weight on the price and little on j's action. Because a trusting j does believe that i is altruistic, she is willing to make sacrifices to persuade her of this trust; on the other hand, a less-trusting j doesn't see much benefit from such costly signaling. Thus, they can separate in equilibrium.

However, this trust-signaling channel is quite weak compared to the classical benefits of an optimal price. In order for this channel to drive an equilibrium where the trusting type forgoes compensation. it is necessary for types to differ sufficiently in their preferences and beliefs. For instance, in Example 2.8, the high type has altruism above $1 - \varepsilon$, while the low type has altruism of less than ε, with $\varepsilon \leq .005$.

In this section, once we have established that it is possible for one player to forgo compensation as a signal of trust, we also use the model to explore whether *both* sides could choose to forgo compensation. In a 1×1 environment, this was not possible; at most one player would want to choose to forgo compensation. It will turn out, again, that here the answer is *yes*, it is now possible.

2.6.1 Setup for 2×2 Environment

This version of the model uses two types for each player. Player 1's types are called $\{L, H\}$ and player 2's are called $\{\ell, h\}$. We assume that $\alpha_L < \alpha_H$ (H is more altruistic) and that $\pi\big(H \mid \ell\big) < \pi\big(H \mid h\big)$ (h is more trusting that 1 is altruistic).[16] We do not take a stand

will be. So, when i trusts that j is the high type, she trusts that j will take actions more in line with i's interests.

16. The assumption $\pi\big(H \mid \ell\big) < \pi\big(H \mid h\big)$ implies that $\pi\big(h \mid L\big) < \pi\big(h \mid H\big)$, i.e. the altruistic 1 (type H) has a higher belief that she is facing a trusting 2 (type h). This is because $\pi\big(H \mid \ell\big) < \pi\big(H \mid h\big) \Leftrightarrow$

on whether $\alpha_\ell \lesseqgtr \alpha_h$, meaning that the more-trusting type h may be either more or less altruistic than the less-trusting ℓ.

In this analysis, we modify the model slightly, in that we now allow either player to choose the price in period 1. As before, one player unilaterally gets to choose the price p, which will then be fixed for trading in period 2; however rather than assume that it is player 1 who gets to choose, we now allow that either player may choose. Formally, assume that a random public draw determines which player choose the price p. We continue to refer to the price chosen by a player of type θ—if she were the one to choose—as p^θ if a pure strategy and as σ^θ if a mixed strategy.[17] Thus, we can refer to both players' strategy of price choice, even though ex post only one of them gets to actualize that choice.

2.6.2 Choosing $p = 0$ as a Signal of Trust

Proposition 2.5 indicates that choosing to forgo compensation cannot in equilibrium signal *altruism* towards one's partner. However, here we show that such a choice can signal *trust* in one's partner, while communicating nothing about one's altruism. Proposition 2.7 demonstrates that there does exist a Perfect Bayesian equilibrium in which there is separation between two types, ℓ and h, who differ in their trust in 1's altruism, with the more trusting type h choosing $p^h = 0$ and the less trusting type ℓ choosing $p^\ell = \frac{1}{2}$:

Proposition 2.7 (2×2 – Signaling trust with $p = 0$)**.**
Let the valuations be distributed $\sim \texttt{Unif}\,[0, 1]$*, and suppose there are two types on each side,* $\{L, H\}$ *for* 1 *and* $\{\ell, h\}$ *for* 2*.*

$\frac{\pi(H,\ell)}{\pi(H,\ell)+\pi(L,\ell)} < \frac{\pi(H,h)}{\pi(H,h)+\pi(L,h)}$ (h being more trusting) requires that $\frac{\pi(H,h)}{\pi(L,h)} > \frac{\pi(H,\ell)}{\pi(L,\ell)}$; rearranging this inequality gives $\frac{\pi(H,h)}{\pi(H,\ell)} > \frac{\pi(L,h)}{\pi(L,\ell)}$ (H has higher belief than L that she is facing h).

17. We make this change because of both notation and content. For notation, this allows us to ask what price the types of 2 would choose; having the model let 2 be a chooser lets us keep referring to the selfish and altruistic types of 1 as L and H, while using ℓ and h to rerer to the less- and more-trusting types of 2, For content, we study in Proposition 2.9 situations where *either player would choose* $p = 0$; for this to be meaningful, we need a model where the price choice is defined for either player.

Then there exists a specification of altruism parameters $\boldsymbol{\alpha}$ *and prior distribution* $\pi(\cdot)$ *in which*

- H *is more altruistic than* L*, i.e.* $\alpha_H > \alpha_L$*, and*
- h *is more trusting than* ℓ*, i.e.* $\pi\big(H \mid h\big) > \pi\big(H \mid \ell\big)$*, but also*
- h *and* ℓ *have the same altruism preferences, i.e.* $\alpha_h = \alpha_\ell$*,*

such that there exists a Perfect Bayesian Equilibrium in which $p^h = 0$ *and* $p^\ell > 0$*.*

The proof appears in Appendix A.2, spelled out in Example 2.8:

Example 2.8 (2×2 Separating Equilibrium)**.**
When parameters $\boldsymbol{\alpha}$ *and type distribution* π *are given by*

$$\begin{aligned} \alpha_L &= 0 \\ \alpha_H &= 1 \\ \alpha_\ell &= 1 \\ \alpha_h &= 1 \end{aligned} \qquad \textit{and} \qquad \begin{bmatrix} \pi(L,\ell) & \pi(L,h) \\ \pi(H,\ell) & \pi(H,h) \end{bmatrix} = \begin{bmatrix} (1-\varepsilon)\,(1-\psi) & \varepsilon\,\psi \\ \varepsilon\,(1-\psi) & (1-\varepsilon)\,\psi \end{bmatrix},$$

for sufficiently small $\varepsilon > 0$ *and for any* $\psi \in (0,1)$*, so that the period-*1 *beliefs held by the two types of* 2 *are*

$$\begin{aligned} \pi\big(H \mid h\big) &= 1 - \varepsilon \\ \pi\big(H \mid \ell\big) &= \varepsilon, \end{aligned}$$

there exists a separating 2×2 *equilibrium in which* 2*'s price choices and* 1*'s updating rules*

are:

$$\begin{array}{ll} p^h = 0 & \\ & \textit{and} \\ p^\ell = \dfrac{1}{2} & \end{array} \qquad \begin{array}{l} \pi\big(h \mid p = 0\big) = 1 \\ \\ \pi\big(h \mid p \neq 0\big) = 0. \end{array}$$

The proof of Proposition 2.7 establishes that h gets a utility boost by persuading 1 of her trust (that $\pi\big(H \mid h\big)$ is high), while ℓ does not get such a boost, and this difference supports the equilibrium in which they select different prices.

To understand the result, consider the the two incentive compatibility conditions that must be met for this for this separating equilibrium to hold. They ensure that neither player wants to deviate to an alternate p'. While ℓ's condition is easy to satisfy, the proof focuses on establishing that for sufficiently small ε, the incentive compatibility condition holds for h: $W^h\big(0 \mid \mathbb{1}_h\big) \geq W^h\big(p' \mid \mathbb{1}_\ell\big)$ for any p'. In other words, the value of the partnership to h is higher when she has high reputation $\mathbb{1}_h$—even at price $p = 0$—than with low reputation ℓ at *any* price p'. Thus h is willing to choose her least favorite price (zero) in order to accrue a reputation for being trusting. Meanwhile, ℓ is near-indifferent to her reputation, since she believes she faces a selfish L who is insensitive to her strategy, so ℓ does not find it worthwhile to sacrifice her favorite price $p = \frac{1}{2}$ to change her reputation.

Figure 6 shows values for the types of 2 in the $\varepsilon > 0$ case, Figure 7 shows the corresponding strategies, and Figure 8 shows values for the extreme $\varepsilon = 0$ case.[18]

The reason players prefer to choose p in this way is that when H believes she is facing a trusting 2 (i.e. of type h), she plays cutoffs κ_b^H and κ_s^H that are more preferable to h. A "more preferable" cutoff is not necessarily a higher one; this depends on the price and on

18. The extreme case $\varepsilon = 0$ is different than the limit of the $\varepsilon > 0$ case. It features a separating equilibrium that does not actually represent signaling, because in this case 1 is 100% certain of 2's type, and hence is not influencable by signaling. Still, this extreme case is worth studying because it satisfies all the same incentive-compatibility inequalities that the small $\varepsilon > 0$ case does, and the algebra is simpler to work through.

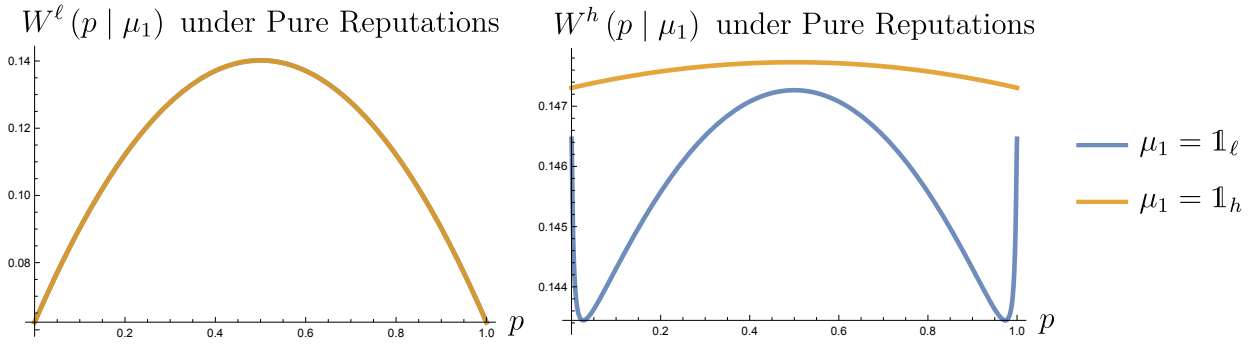

Figure 6: The value of the partnership to each type of 2, given parameters from Example 2.8 with $\varepsilon = .005$, under two possible reputations that 1 could hold: either $\mathbb{1}_h$ where 1 is sure 2 is h, or $\mathbb{1}_\ell$ where 1 is sure 2 is ℓ. They show why the separating 2×2 equilibrium from Example 2.8 is Incentive Compatible. In the equilibrium, 1 infers that 2 is h if she observes her choose price $p = 0$, and believes that 2 is ℓ otherwise. Facing this response, ℓ chooses $p^\ell = \frac{1}{2}$, because she gets highest utility at $p = \frac{1}{2}$ on either curve, and because she does not care about influencing her reputation. Indeed, she gets nearly the same utility for either reputation (in the W^ℓ figure on the left, the blue $W^\ell(\cdot \mid \mathbb{1}_\ell)$ curve cannot even be seen because it is covered up by the $W^\ell(\cdot \mid \mathbb{1}_h)$ curve on the right). On the other hand, h choose $p^h = 0$, because even though on either curve she gets the highest payoff at $p = \frac{1}{2}$, she does better with a high reputation at 0, getting utility $W^h(0 \mid \mathbb{1}_h)$, than with a low reputation anywhere else, getting at best $W^\ell\left(\frac{1}{2} \mid \mathbb{1}_\ell\right)$.

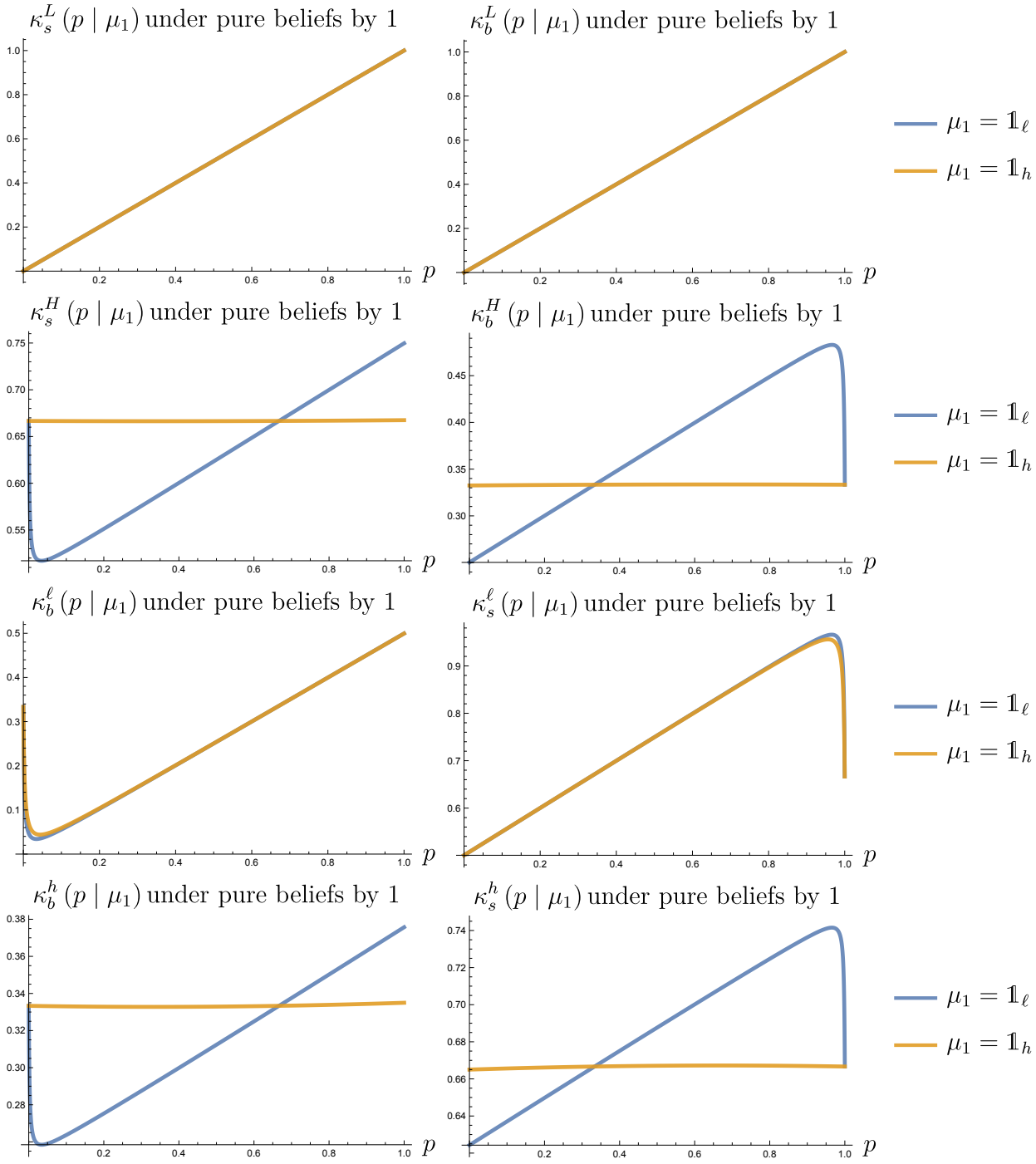

Figure 7: Strategies by all types, given parameters from Example 2.8 with $\varepsilon = .005$, under both role assignments, and under the two possible reputations that 1 could hold: either $\mathbb{1}_h$ where 1 is sure 2 is h, or $\mathbb{1}_\ell$ where 1 is sure 2 is ℓ. They show that strategies are close to linear in p—as they would be at $\varepsilon = 0$ and as depicted in footnote 22—except at p close to 0 or 1, where ℓ infers that she likely is facing H, not L as her prior led her to believe. These are the strategies that underlie the value of the partnership shown in Figure 6.

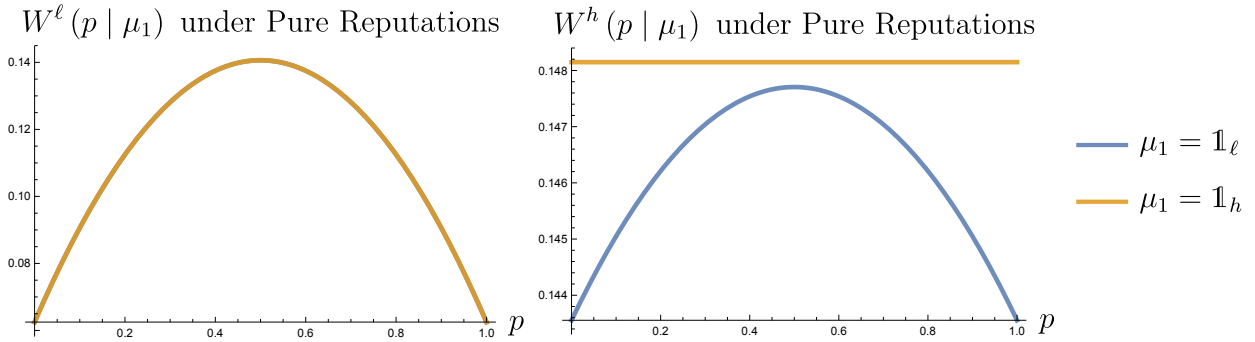

Figure 8: The value of the partnership to each type of 2, given parameters from Example 2.8 with $\varepsilon = 0$, under two possible reputations that 1 could hold: either $\mathbb{1}_h$ where 1 is sure 2 is h, or $\mathbb{1}_\ell$ where 1 is sure 2 is ℓ. Here, the separating 2×2 equilibrium also obtains, though with h totally indifferent to price because she believes that her parter is H, who is indifferent to the size of a transfer between the two players, just like she is. As in the $\varepsilon > 0$ case, h does better with a high reputation $\mathbb{1}$ than a low one, though here her indifference means that it would be no sacrifice to choose $p = 0$ over any other. Moreover, in no equilibrium could her partner H adopt a belief other than $\mathbb{1}_H$, so h's price choice does not affect her reputation in any case.

one's own cutoff. To see this laid out, take for example the role assignment with 1 as the buyer and 2 the seller, in the case of Example 2.8 with $\varepsilon = 0$ (where h is certain that 1's type is H, and where $\alpha_h = 1$).[19] In this case, h's ex ante utility (from (2.5) and (1.11)) is

$$\mathcal{U}_s^h\Big(p \mid \kappa_b^H, \kappa_s^h\Big) = \left(-\frac{0 + \kappa_s^h}{2} + 1 \cdot \frac{1 + \kappa_b^H}{2}\right) \cdot (1 - \kappa_b^H) \cdot \kappa_s^h.$$

Type h's preferred value of her partner's cutoff κ_b^H maximizes this utility (given that h chooses κ_s^h to best-respond to it), and to do so it must balance two channels. First, on the intensive margin, a higher κ_b^H means a higher average buyer valuation conditional on buying (via increasing the term $\frac{1+\kappa_b^H}{2} = \mathbb{E}\Big[v_b \mid v_b \geq \kappa_b^H\Big]$), which h likes. But second, on the extensive margin, a higher κ_b^H entails fewer transactions taking place (via decreasing

19. While this case illustrates equilibrium motivations succinctly, these extreme believfs cannot give rise to an equilibrium with signaling. This is because, given the common prior assumption, if $\pi\big(H \mid h\big) = \pi\big(L \mid \ell\big) = 1$, then the prior probability must include $\pi(H, \ell) = 0$, which in turn implies that $\pi\big(h \mid H\big) = 1$, so there H's beliefs cannot be altered by any new information.

term $(1-\kappa_b^H) = 1 - F_b\Big(\kappa_b^H\Big)$), which h dislikes.[20]

It turns out that under these parameters, h's preferred partner cutoff to face is $\kappa_b^H = \frac{1}{3}$, and because their preferences are completely aligned ($\alpha_h = \alpha_H = 1$), this is just what H chooses in equilibrium when she believes she is facing h.[21] On the other hand, when H believes she is facing ℓ, she best-responds to ℓ's strategy of $\kappa_s^\ell = \frac{1+p}{2}$ by playing a different strategy: $\kappa_b^H = \frac{1+p}{4}$.[22] For almost every p, this strategy κ_b^H is less preferable to h: if $p < \frac{1}{3}$, this κ_b^H is too low (H asks too many low-value favors), and if $p > \frac{1}{3}$ it is too high (H is too reticent to ask for high-value favors). Thus, h prefers for H believe that 1's type is the h, the trusting one.

In summary, when h knows she is facing H, both her interests and her information are aligned with those of h. That makes her action more amenable to h, and thus h prefers reputation $\mathbb{1}_h$.[23]

---

20. To be more explicit, h dislikes it when fewer transactions take place so long as she gets positive utility from the average transaction, that is, if $-\frac{0+\kappa_s^h}{2} + \frac{1+\kappa_b^H}{2} \geq 0$. This inequality always holds in equilibrium, because every player's first order condition is that they be indifferent to a trade at their partner's average valuation and their own marginal one (cutoff), so they strictly benefit from a trade substituting their own average valuation for their marginal one.

21. This equilibrium strategy under $\mu_H = \mathbb{1}_h$ comes from solving the first-order conditions where H and h best-respond to one another only:

$$\begin{aligned} \kappa_b^H &= \frac{0+\kappa_s^h}{2} & \kappa_s^h &= \frac{1+\kappa_b^H}{2} \\ \implies \kappa_b^H &= \frac{1}{3} & \kappa_s^h &= \frac{2}{3}. \end{aligned}$$

22. The equilibrium under $\mu_H = \mu_L = \mathbb{1}_\ell$ and $\varepsilon = 0$ comes from solving the four first-order conditions. This can be done sequentially: (1) L is selfish, so her strategy does not depend on anyone else's, then (2) ℓ best-responds to only L, then (3) H best-responds to ℓ, and finally (4) h best-responds to H. They come out as follows:

$$\begin{aligned} \kappa_b^L &= 0 \cdot \frac{0+\kappa_s^\ell}{2} + 1 \cdot p & \kappa_s^\ell &= 1 \cdot \frac{1+\kappa_b^L}{2} + 0 \cdot p & \kappa_b^H &= 1 \cdot \frac{0+\kappa_s^\ell}{2} + 0 \cdot p & \kappa_s^h &= 1 \cdot \frac{1+\kappa_b^H}{2} + 0 \cdot p \\ \implies \kappa_b^L &= p & \kappa_s^\ell &= \frac{1+p}{2} & \kappa_b^H &= \frac{0+\frac{1+p}{2}}{2} = \frac{1+p}{4} & \kappa_s^h &= \frac{1+\frac{1+p}{4}}{2} = \frac{5+p}{8} \end{aligned} \tag{2.13}$$

23. With parameters other than these totally altruistic preferences, there is in general a wedge between the two players' preferred strategy. Then, each player i prefers her partner to employ a higher or lower cutoff

2.6.3 Mutually choosing $p = 0$ as a Signal of Trust

Thus far, any equilibrium we have seen that involves someone forgoing compensation has featured just one individual making this choice; her partner, if given the option, would instead prefer to set a price higher than zero. However, in the friendship situation we seek to model, the two friends do not disagree on whether to forgo compensation. In fact, in a 2×2 context, there *do* exist equilibria in which both sides would choose to forgo money to signal their type. This can occur when both sides feature a trusting type who is also altruistic. Each type's equilibrium price is defined as what they would choose, if they were selected to name the price.

Proposition 2.9 (2×2 – Symmetrically signaling trust with $p = 0$)**.**
Let the valuations be unit-uniformly distributed as $v_i \sim \texttt{Unif}\,[0, 1]$, and suppose there are two types on each side, $\{L, H\}$ for 1 and $\{\ell, h\}$ for 2.

Then there exists a specification of altruism parameters $\boldsymbol{\alpha}$ and prior distribution π in which

- *H and h are more altruistic than L and ℓ, i.e. $\alpha_\ell < \alpha_h$ and $\alpha_L < \alpha_H$, and*
- *H and h are more trusting than L and ℓ, i.e. $\pi\big(H \mid \ell\big) < \pi\big(H \mid h\big)$ and $\pi\big(h \mid L\big) < \pi\big(h \mid H\big)$,*

such that there exists a Perfect Bayesian Equilibrium in which if 1 chooses the price, then $p^L = 0$ and $p^H > 0$, while if 2 chooses the price, then $p^\ell = 0$ and $p^h > 0$.

The proof appears in Appendix A.2.

The proof is by construction. Example 2.10 lays out the parameters, equilibrium strategies and updating rules, and Figure 9 shows the payoffs the players face given possible high and

depending on the price p, putting less weight on j's transfer and valuation, and more on i's transfer and valuation (both of opposite sign to j's).

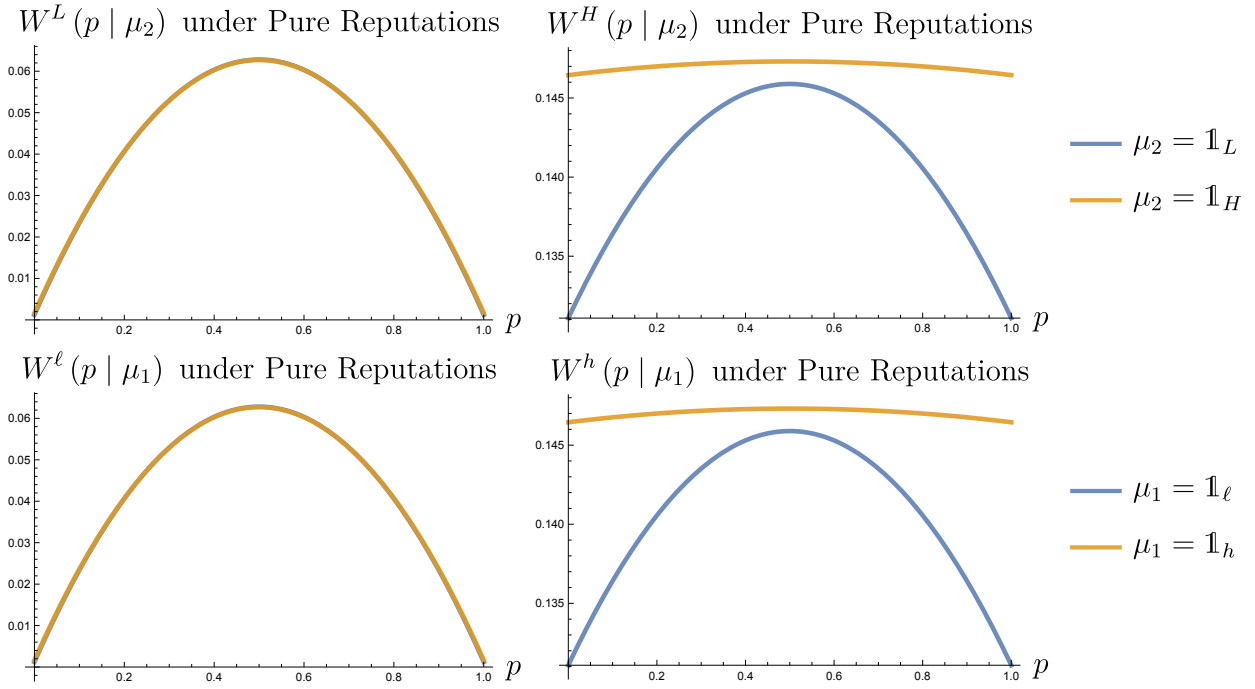

Figure 9: Payoffs associated with parameters from Example 2.10. The top two show payoffs to L and H, the two types of 1, given low and high reputations; likewise, the bottom two show payoffs to ℓ and h, the two types of 2, given low and high reputations. They establish that equilibrium strategies put forth in Example 2.10 are Incentive Compatible. For instance, the top two show that H would prefer to choose $p = 0$ and garner a high reputation, rather than choose $p = \frac{1}{2}$ and garner a low reputation.

low reputations. This example features symmetric parameters, where swapping the players' names does not change their altruism or beliefs (e.g. $\alpha_H = \alpha_h$), though this symmetry is not strictly necessary for the result; adjusting parameters slightly would not change equilibrium existence. As with Proposition 2.7, separation is Incentive Compatible because the high type benefits enough from a high reputation to make the sacrifice of an more-efficient price worthwhile, while the low type barely benefits at all from a high reputation since she believes she faces a selfish and pessimistic partner.

Example 2.10 (2×2 Symmetric Separating Equilibrium)**.**

With (symmetric) parameter values

$$\begin{array}{c} \alpha_L = \alpha_\ell = .01 \\ \alpha_H = \alpha_h = .99 \end{array} \qquad \textit{and} \qquad \begin{bmatrix} \pi(L,\ell) & \pi(L,h) \\ \pi(H,\ell) & \pi(H,h) \end{bmatrix} = \begin{bmatrix} .495 & .005 \\ .005 & .495 \end{bmatrix},$$

*so that the period-*1 *beliefs of* 1 *if* 2 *chooses price, and of* 2 *if* 1 *chooses price, are*

$$\pi\big(H \mid h\big) = \pi\big(h \mid h\big) = .99$$
$$\pi\big(H \mid \ell\big) = \pi\big(h \mid L\big) = .01$$

there exists a pooling 2×2 *equilibrium in which the two players' price choices (if they get to choose the price), and their updating rules (if their partner does) are:*

$$\begin{array}{c} p^H = p^h = 0 \\ p^L = p^\ell = \dfrac{1}{2} \end{array} \qquad \textit{and} \qquad \begin{array}{l} \pi\big(H \mid p = 0\big) = \pi\big(h \mid p = 0\big) = 1 \\ \pi\big(H \mid p \neq 0\big) = \pi\big(h \mid p \neq 0\big) = 0. \end{array}$$

In the separating equilibrium the high type selects $p = 0$ because she gains by persuading her partner that she is both altruistic and trusting. As with Example 2.8, the high type, believing she is facing an altruist, gains from such a reputation, and this reputational gain is larger than the loss from not selecting $p = \frac{1}{2}$, which would be her preferred price if her reputation were fixed.

2.6.4 Putting the Trust-Signaling Results in Context

Proposition 2.7 establishes the existence of a separating equilibrium where the trusting type signals her type by selecting $p = 0$, to earn a reputation as trusting. However, to understand it in context and not overstate the findings of the model, two examples illustrate what else could happen in this situation. They show that while the model does support the

interpretation of forgoing compensation as trust-signaling, it is not the only possible outcome of the model.

The fist example emphasizes that the model relies on extreme parameters for its conclusion, but when those parameters are relaxed, the separating equilibrium no longer holds. Example 2.8 involved a large difference in altruism between the two types of 1, with $\alpha_H = 1 - \varepsilon$ and $\alpha_L = \varepsilon$, as well as a large difference between beliefs held by 2, with $\pi\big(H \mid h\big) = 1 - \varepsilon$ and $\pi\big(H \mid \ell\big) = \varepsilon$, both assuming $\varepsilon < .005$.

Example 2.11, and Figure 10, show how with a slightly larger ε, the separating equilibrium of Example 2.8 breaks down:

Example 2.11 (2×2 Lack of Separating Equilibrium)**.**
With parameter values like those of Example 2.8 but with $\varepsilon = .01$*, namely*

$$\begin{aligned} \alpha_L &= 0 \\ \alpha_H &= .99 \\ \alpha_\ell &= .99 \\ \alpha_h &= .99 \end{aligned} \qquad \textit{and} \qquad \begin{bmatrix} \pi(L,\ell) & \pi(L,h) \\ \pi(H,\ell) & \pi(H,h) \end{bmatrix} = \begin{bmatrix} .99\,(1-\psi) & .01\,\psi \\ .01\,(1-\psi) & .99\,\psi \end{bmatrix},$$

for any $\psi \in (0,1)$*, so that the period-*1 *beliefs of* 2 *are*

$$\begin{aligned} \pi\big(H \mid h\big) &= .99 \\ \pi\big(H \mid \ell\big) &= .01, \end{aligned}$$

there the following profile of 2*'s price choices and* 1*'s updating rules does* not *constitute a*

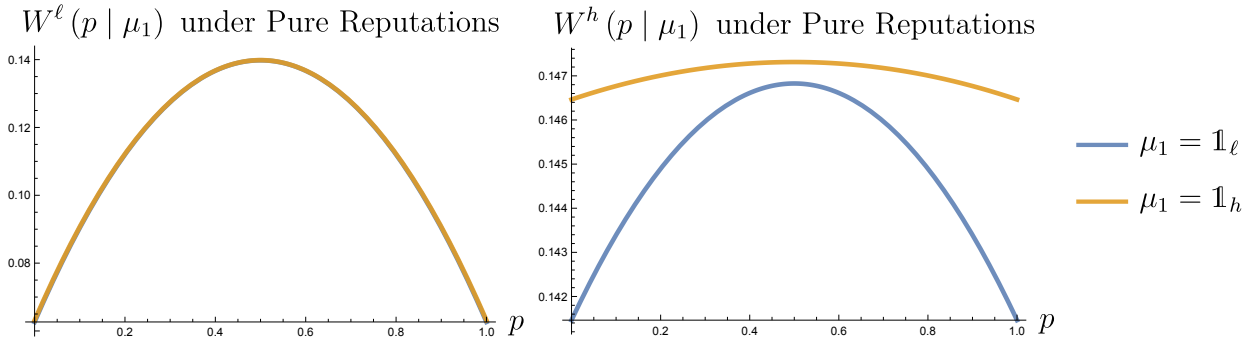

Figure 10: The value of the partnership to each type of 2, given parameters from Example 2.11 (where $\varepsilon = .01$), under two possible reputations that 1 could hold: either $\mathbb{1}_h$ where 1 is sure 2's type is h, or $\mathbb{1}_\ell$ where 1 is sure 2's type is ℓ. They show why the parameters from Example 2.11 do not support a separating equilibrium. If there were a strategy and belief profile where h was supposed to play $p^h = 0$ and ℓ was supposed to play $p^\ell = \frac{1}{2}$, then h would do better to deviate and mimic ℓ, rather than to stick with her prescribed action, since here $W^h\left(\frac{1}{2} \mid \mathbb{1}_\ell\right) > W^h\big(0 \mid \mathbb{1}_h\big)$.

separating equilibrium:

$$\begin{array}{ll} p^h = 0 & \pi\big(h \mid p = 0\big) = 1 \\ \quad \textit{and} & \\ p^\ell = \dfrac{1}{2} & \pi\big(h \mid p \neq 0\big) = 0. \end{array}$$

The issue is that under these parameter values, it is no longer incentive compatible for h to choose $p = 0$, even if this garnered reputation $\mathbb{1}_h$. As Figure 10 illustrates, while $W^h\big(0 \mid \mathbb{1}_h\big) \not> W^h\big(0 \mid \mathbb{1}_\ell\big)$, meaning the reputation still has positive value, now $W^h\big(0 \mid \mathbb{1}_h\big) \not> W^h\left(\frac{1}{2} \mid \mathbb{1}_\ell\right)$, so it is not enough value to keep h from preferring to deviate away from a price of zero to a price of $\frac{1}{2}$.

The reason for this change, given the larger value of ε here than Example 2.8, is that the value of a trusting reputation comes from strategic complementarity in cutoff strategies: when 1 believes that 2 is trusting, then 1 plays a higher strategy because she believes that 2 is playing one as well. However, this effect is mediated by their altruism parameters which govern how strongly each player's best-response depends on her partner's; as these altruism parameters get further from 1, the effect is diminished in magnitude.

For the second example, we show that separating is not the only Perfect Bayesian Equilibrium. Even with the same parameter values as Example 2.8, there also exists a pooling equilibrium, in which both types of 2 select the same price, and consequently 1 does no updating of her prior about 2's type, no matter what price she sees 2 choose:

Example 2.12 (2×2 Pooling Equilibrium)**.**

With a particular case of the same parameter values as Example 2.6 (now setting $\varepsilon = .005$*),*

$$\begin{aligned} \alpha_L &= 0 \\ \alpha_H &= .995 \\ \alpha_\ell &= .995 \\ \alpha_h &= .995 \end{aligned} \qquad \textit{and} \qquad \begin{bmatrix} \pi(L,\ell) & \pi(L,h) \\ \pi(H,\ell) & \pi(H,h) \end{bmatrix} = \begin{bmatrix} .995 \cdot (1-\psi) & .005 \cdot \psi \\ .005 \cdot (1-\psi) & .995 \cdot \psi \end{bmatrix},$$

for any $\psi \in (0,1)$*, so that the period-*1 *beliefs of* 2 *are*

$$\begin{aligned} \pi\big(H \mid h\big) &= .995 \\ \pi\big(H \mid \ell\big) &= .005, \end{aligned}$$

there exists a pooling 2×2 *equilibrium in which* 2*'s price choices and* 1*'s updating rules are:*

$$\begin{aligned} p^h &= \frac{1}{2} \\ p^\ell &= \frac{1}{2} \end{aligned} \qquad \textit{and} \quad \pi\big(h \mid p\big) = \frac{1}{2} \quad \forall p.$$

Figure 11 illustrates the payoffs associated with this pooling equilibrium. Since 1 will stick to her prior $\pi(h) = \frac{1}{2}$ no matter what p she observers, both ℓ and h prefer to stick to $p = \frac{1}{2}$, as it gives the highest payoff at that prior reputation. In equilibrium, Bayes' rule vindicates 1's updating rule at $p = \frac{1}{2}$, and imposes no restrictions for other p since those prices are never chosen by 2 in equilibrium.

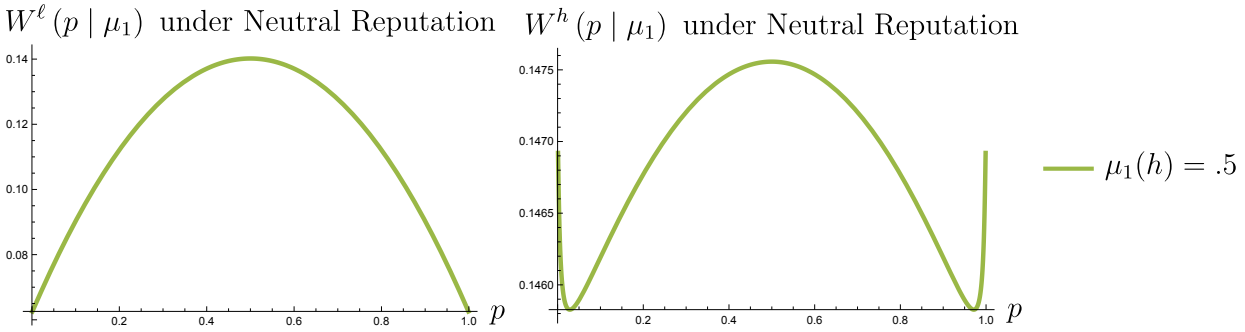

Figure 11: The value of the partnership to each type of 2, under a reputation that puts equal weight on ℓ and h, for parameters in Examples 2.8 and 2.12. This pooling equilibrium is Incentive Compatible because, given that 1 maintains her prior belief $\pi(h) = \pi(\ell) = \frac{1}{2}$, both types of 2 do best by playing $p^h = p^\ell = \frac{1}{2}$, as the figures show.

This pooling equilibrium leads to slightly higher payoffs to h than does the separating one, which suggests that the signaling is inefficient. At least under these parameters, it would be better if it weren't necessary to sacrifice a more efficient price just to prove one's trust. However, if 1 expects this behavior, then it is Incentive Compatible for h to comply. The model therefore implies that if two friends were stuck in an equilibrium where they feel the need to forswear compensation to prove themselves, that they could do better if they could switch to an equilibrium where no one made inferences based on anyone's stance towards the price for help between friends.

CHAPTER 3

TOWARD A NEW MODEL OF FRIENDSHIP

3.1 Extensions and Further Models of Friendship

The model developed in these chapter provided one view of when friends would—and would not—decide that the services they provide one another will be for free. The model indicates that except in special circumstances, they would prefer to use money. Still, these conclusions are derived in a specific environment, and because this, it would be useful to explore how robust these conclusions are to extensions of the model. Further, the model, or one like it, could be used to explore related questions. Here I describe several of the most valuable potential extensions and new directions, and outline their prospects and difficulties.

Embedding in a Broader Market Environment

To focus on the incentives that arise due to trade itself, this model treats the two people as exogenously stuck in a partnership. The participants never face a choice of whether to commence with the partnership at all; they simply choose the terms of trade. Because they are always free to refuse a favor, and because the outside option is 0, this assumption is unproblematic here: they could replicate exiting by simply refusing to trade.

However, in real life, maintaining friendships takes work, and exit is an option. Chassang (2010) explores a model where uncertainty about one's partner's motives can lead a relationship to unravel and the partners ultimately to exit; this model is driven by (built-in) strategic complementarity, so if one party contributes less, the other may do so as well.

Including exit in this model might make the value of an altruistic reputation higher, even at prices like $p = \frac{1}{2}$ where in this model it tended to be negative. The reason is that it

could become useful to be seen as altruistic to keep friends around. If the model included an additional value of an altruistic or trusting reputation, then it might provide a stronger signaling motive to forgo money. However, the results here suggest that this model might still not explain forgoing money at reasonable parameters. A successful model would need to explain not just why it is useful to be seen as altruistic, but why bona fide altruists get more value from it; otherwise a separating equilibrium cannot be sustained, as the selfish would also seek to cultivate an altruistic reputation. Proposition 2.5 suggests that, so long as the reason to be seen as altruistic is to get more out of one's trading partners, then selfish people have the strongest incentive to be seen as altruistic, which would unravel any such equilibrium.

Extending the Domain P

One assumption begging to be relaxed is that price is restricted to domain $p \in P = [0, 1]$. This restriction was set primarily for tractability, to ensure that there is zero probability that the price will be outside the support of the buyer's and seller's valuations $v_{\mathtt{b}i}$ and $v_{\mathtt{s}j}$, which are distributed $\mathtt{Unif}\,[0, 1]$. In a model without altruism, this assumption would not be an issue – selecting such an extreme price would be tantamount to choosing autarky, since either the buyer or the seller would be guaranteed to get negative material utility. However, under altruistic preferences, this is no longer the case, as individuals may derive positive utility even if they get negative material utility, because they internalize some of their friend's material utility. The reason that restricting $p \in [0, 1]$ helps with tractability is because, in Proposition 1.1, it guarantees that the best-response cutoff κ_i^* is itself in $[0, 1]$.

If p were unrestricted in $\mathbb{R}$, then equations (1.5)–(1.6) would need to be modified by clipping κ_i^* to $[0, 1]$, i.e. by applying the function $\max\big\{0, \min\{1, \cdot\}\big\}$ to the RHS. This change would complicate solving for equilibrium and conducting comparative statics, though it would not make the task insurmountable.

What would be the likely result of this analysis? First, while Proposition 1.11 on the marginal value of compensation (defined as $W'\Big(0 \mid \alpha_i, \alpha_j\Big)$) would not be changed, Proposition 2.4 on the equilibrium choice of p in the 1×1 game, would no longer be valid. Being a local max would no longer imply being a local max. At $p = 0$ the best-response formulae, and hence the equilibrium and the value of the partnership, would be continuous, so the left and right derivatives of the value would be equal. Thus, if conditions were such that i preferred $p = 0$ over $p = \frac{1}{2}$, then a small negative p would be even better.

This analysis demonstrates that there is nothing intrinsically special about a price of $p = 0$ in the 1×1 game. If i takes advantage of j' altruism, she does so by choosing an *extreme* price, not so much a zero one. Either she chooses a low price (possibly negative, if available), under which she herself will rarely be selling, and so she will benefit from this low price as a buyer, or conversely she chooses a high price where she benefits as a seller.

An unrestricted domain for p would not change the 2×1 results in Section 2.5 that, a price of zero cannot be a signal of altruism. If a particular equilibrium does not exist, because a profitable deviation from that choice is available, then expanding the domain would not change this fact.

But the impact on the 2×2 trust-signaling results of Section 2.6 remains uncertain. These results rely on the value of reputation being so high for the high type h that an equilibrium exists where it is not worthwhile to select a more efficient price than $p = 0$ if it means giving up this reputation. With an unrestricted domain, it is possible that a very extreme p is more attractive—even at an unfavorable reputation—unraveling the equilibrium. While this remains an open question, it is not an unsolvable problem; the equilibrium can still be calculated just as well, it only takes more cases to work through.

Generalizing the Role Probabilities

In this model, when we allowed players to take on either role of buyer or seller, we did so by assuming that the two were drawn with equal probability. This assumption was both substantive, to model a friendship between two equals, and tactical, to make use of symmetry to prove results.

However, real life situations could differ from this situation in two ways.

First, the role probabilities might be unbalanced. One person in the relationship would be more likely to be in a position to provide help, and the other to need help, such as as a senior and junior colleague. This change would tend to amplify the classical aspects of compensation's usefulness, and push the likely buyer to want a low price and the likely seller to want a high one. (However, the buyer would not necessarily want $p = 0$, because as an optimal monopsonist, she wants the seller to have some motivation to sell.)

Second, each individual's role probability might be private information.[1] In this case, it is plausible that one's private information would influence her choice of price p. However, there doesn't seem to be a signaling rationale for this. There is no strategic rationale for (say) i to convince j that i is likely to be a seller, because once j has to take an action in the bilateral trading game (namely, deciding whether or not to trade), j will have already observed the role realizations.

Adjusting the Fixed-Price Mechanism

In the model, we assumed that in the period 2 bilateral trading game, there are only two available actions—to agree or to disagree to trade. We also assumed that in period 1, the

1. There is no reason the role probabilities of the two need sum to 1 – e.g. it may be that one needs a favor (i.e. is the buyer) but the other is utterly unable to provide it. In this way, knowing one's own role assignment need not provide information about one's partner's.

only choices were to pick a single real number p. These assumptions restrict the players in two ways.

First, the messaging space is very coarse. In this model, these friends care about one another's costs and benefits. In in real life, friends communicate not just whether their their request is urgent enough to ask, and if their circumstances are too difficult to comply, but also other information about their situation. For instance, when declining to provide help, we usually give an excuse for why we don't have the time or are why our help won't actually be useful, effectively arguing that our costs are high or the benefits low. (When we do this we provide a rationale for why the trade is not so efficient, rather than admitting when we simply do not care enough about the asker's problem for it to be worth our while to help.) So it makes sense to have more messages available than "yes" and "no". Bhaskar and Sadler (2017) build a model that pursues this avenue. There, two individuals with altruistic preferences (or equivalently, with consumption externalities) play a variant of this bilateral trading with $p = 0$, and choose which message to convey. The paper focuses on (and proves the optimality of) "hierarchical mechanisms", in which the messages correspond to different intervals of valuation draws. Welfare is increasing in the number of messages chosen in equilibrium; this immediately implies that our mechanism with two messages is not the optimal one.[2]

The second restriction here is that the mechanism only lets the two individuals choose a single price. This restriction was adopted chiefly for tractability, to draw attention towards the strategic complementarity in generosity cutoffs and to the value of reputation. However, even Chatterjee and Samuelson (1983), introducing and solving the bilateral trading game (with selfish players), solved the game with strategies where price offers are increasing in

2. Their model also differs from ours in that neither player has property rights over the good or service: if both say they value it highly, then a coin flip determines who gets it. In contrast, this paper's bilateral trading model gives the seller property rights (i.e. veto power): if the buyer wants the service but the seller doesn't want the buyer to have it, the seller "keeps" it, and no trade takes place. Interestingly, in the two-message case (which this present paper explores), this assumption of equal property rights totally eliminates

valuation.[3] And furthermore, Maskin and Tirole (1990, 1992) describe mechanism design where an informed principal chooses the mechanism from a very general set. So there is ample ground to study richer environments of prices between friends.

The rub is solving the thing. Although in a 1×1 environment (where altruism parameters are common knowledge), it is straightforward to solve the Chatterjee and Samuelson (1983) bilateral bargaining game with altruism and with valuations drawn from $\texttt{Unif}\,[0,1]$, once there are multiple types, solving for equilibrium is not so simple. While the first-order conditions suggest that linear strategies will solve equilibrium, they only do so locally. When solving for the optimal inverse strategies as a function of price (what valuation a type must have, to name a particular price), at a high enough price some types reach the end of their valuation support, which make the posterior belief and the best-response of the "remaining" types jump discontinuously. Patching these regions together does not seem to yield a solution. I therefore conjecture that with multiple altruism types, this bilateral bargaining game in general does not have an equilibrium in linear strategies. Characterizing the full set of equilibria is an open question.

Nonetheless, a fixed price p is indeed one equilibrium of the original Chatterjee and Samuelson (1983) game. In addition, if in period 1, one player conveys that she will play a p-fixed-price strategy, then it is optimal for her partner to do the same. This justifies the restriction to fixed-price mechanism as something endogenous to the model, rather than an exogenous restriction to the game. (However, if we took this view seriously, we would have to properly specify the set of mechanisms from which agents could choose, since in the period-1 signaling game, agents might be tempted to deviate to some non-fixed-price mechanism, and her partner would in equilibrium infer from this choice something about her type.)

---

the strategic aspects: the best-response cutoff no longer depends on the opponent's choice of cutoff.

3. See also Kucuksenel (2012) on extending some classic mechanism design to players with social preferences, albeit equal and commonly known preferences.

Lengthening to a Repeated Trading Game

Real friends help each other out over time, not just in a one-shot bilateral trading game. For this model, I selected a static environment because reputation plays an important role, yet dynamic two-sided reputation models are notoriously difficult to solve (Mailath and Samuelson, 2015). Since the main insights could be captured in a static model, there was no need for these extra complications. Still, extending the model to multiple periods would be interesting.

It is an open question if the mechanisms of this model would survive to a dynamic environment, for two reasons. First, Proposition 2.7, on signaling trust in a 2×2 information environment, requires the value of a high reputation (i.e. that you trust in your friend's altruism) being high. In a repeated game, the value of reputation may be lower, since after observing each other's trading behavior, players would start to infer their type anyway. For example, in a related model of favor trading based on Abdulkadiroğlu and Bagwell (2013), Kalla (2010) shows that in a simple equilibrium, patient types separate from impatient types at their first opportunity for providing a favor. If you will soon learn your friend's type, then your initial perception of her is unlikely to matter much to the friendship, so signaling is not as valuable to her.

The second reason a dynamic model may differ from a static one is that, even with $p = 0$, the opportunity to trade favors dynamically may partially replicate the impact of having prices in the static model. This could make a reputation for altruism may be less valuable than one for selfishness, as can occur at $p = \frac{1}{2}$ in this model. Just as when payment is in money, facing a more altruistic partner means that a selfish i can take advantage of her kindness by striking a harder bargain. If i thinks that j will be quite generous to her regardless of whether she is generous today, then i will have a low incentive to do a favor for j. In fact, i has diminishing returns in j's generosity cutoff κ_j, because the most valuable favors to a

generous i are those that are low-cost to j. Thus, altruism introduces a natural concavity into the returns to generosity in favor-trading models.

Adding Friction to the Means of Exchange

In this model, payment is frictionless: a transfer at price p has the same cost to the payer as it has benefit to the recipient; the only choice is what size that transfer should be. This means that the model cannot address the *means* of payback, only the size. An extension of this model could help explore when friends pay each other back in some other less-direct way.

There are two types of payback worth exploring. First, the "buyer" of the service may pay back the "seller" by being more amenable to providing a service in the future. As discussed in the dynamics section above, this paper's model could be extended to such a repeated game, though the reputation dynamics would be difficult to analyze. (Such a dynamic is at the heart of papers like Bandiera et al. (2005), exploring whether seemingly generous behavior is in reality just a selfish response to dynamic incentives.)

The second possible generalization of payback involves a lossy transfer. This generalization is easy to add to the model, while keeping it static. In period 1, the person choosing the rules for period-2 bilateral trading game would pick not only a price $p \in P$, but also a wedge $\tau \in T \subset \mathbb{R}$. The buyer would pay p, but the seller would receive only $(1-\tau)p$. For instance, a wedge set $T = \{0, .13\}$ would represent a situation where the two friends have two choices – they could either pre-agree that transfers would use direct with money ($\tau = 0$), or they could pre-agree that transfers would be in-kind with goods with some deadweight loss on average ($\tau = .13$). I expect that a lossy transfer would function similarly to an inefficient price, so that the analysis would reveal that in a "2×1" environment like Section 2.5, high-altruism types would choose the more efficient means of transfer, while in a "2×2" environment,

choosing an inefficient means of transfer could, in equilibrium, serve as a signal of altruism and trust, if the returns from separating were high enough. Because introducing such a τ would keep the best-response equations (1.5)–(1.6) and (2.7)–(2.8) linear in p, it would not make the equilibrium any harder to solve.

In this way, the model could be used to investigate possible signaling motivations using for in-kind payment, driven by a purely instrumental preference for reputation.

Broadening the Valuation Distributions

One aspect of friendship missing from this model is that it is not merely transactional, as Sandel (2012) likes to remind us. Friends often do activities together that they both enjoy. In the model as written, every transaction involves a cost to one party and a benefit to the other (both buyer valuation v_b and seller valuation v_s are ≥ 0). The model was built this way built for both concordance with existing bilateral trading models, and because this makes it unambiguous who is the buyer and who is the seller. This distinction was important because this paper used a very simple mechanism: a fixed price p, paid from the buyer to the seller (and it is common knowledge which plays which role), so there is never ambiguity about who is supposed to pay whom.

Generalizing the valuation distribution to allow negative costs would let us use the model to analyze questions such as: when friends go bowling, why doesn't the more enthusiastic friend pay the more reluctant one to come out? Instead, this payment may take the form of decision rights over what activity to do next time, rather than money; why is this?

Other ways of generalizing the valuation distribution involve correlation and information. For instance, friends sometimes pay each other for the component of their cost that is monetary and easy to measure. For instance, might pay or reimburse for gas and snacks on a road trip (but not pay for hassle or wear and tear on the driver's car), or pay back for movie tickets (but

not for the hassle of securing the tickets or organizing the outing). This could be modeled as a cost to the seller that is partly common knowledge, and partly uncertain; the common-knowledge component is the monetary costs incurred, while the uncertain component is the hassle or depreciation to long-lived goods. In the model as presented here, the common knowledge component was 0, with costs in $[0, 1]$ added on top. However, the same analysis of why the price is 0 might be applied to these other situations, to understand why the explicit price is "those expenses that are measured in money".

A second way the model could be generalized is to more directly explore different valuation distributions, rather than assume that all valuations are distributed $\overset{\texttt{iid}}{\sim} \texttt{Unif}\,[0, 1]$. This assumption was, again, introduced for tractability – it makes the best-response functions linear, and so it makes equilibrium easier to solve. For instance, Bhaskar and Sadler (2017) prove various properties of a related model, with generalized valuation distributions. Not only would this make the model more flexible, but it might introduce new dimensions of interest. For instance, even a linear distribution introduces concavity into the value of the trading game, as a function of the partner's cutoff. (There are diminishing returns because the seller performs the easy favors first.) It is possible that this concavity (possibly stronger with some other distribution of cost or benefit valuations), combined with strategic complementarity, could lead to a preference for ignorance. In this situation, i would rather not learn more information about j's type (assuming that when i learned information, this fact would be common knowledge). In the spirit of Dalkiran et al. (2012) and Roy (2012), friends might prefer to leave it vague just how they feel about one another. Avoiding speaking of prices would allow them to remain strategically ignorant, whereas if there were some payment, there would have to be finer information about the price chosen.

3.2 Conclusions

This paper studied a model of friendship as a bilateral trading relationship between altruists. Although the model captures only certain aspects of friendship, it has taught us several things.

First, in a friendship, generosity cutoffs are strategic complements (Proposition 1.5). The model naturally gives rise to responding to kindness with kindness, without assuming a direct preference for reciprocity as in models of interdependent preferences (Sobel, 2005; Falk and Fischbacher, 2006; Gul and Pesendorfer, 2010)). Due to this strategic complementarity, friends perform and request an inefficient quantity of favors; they would do better by committing to be more generous.

Second when the two parties know each other's preferences well, then the model does not provide much support for friends choosing to forgo money (Propositions 1.10–1.11 and Corollary 1.13). Rather, someone who cares more for her friend has a greater incentive to choose a more efficient price (here, $p = \frac{1}{2}$). Only if one party is much more selfish than the other would she prefer to operate for free, in order to mooch off her "friend" by asking a lot of favors and not paying for them.

Third, the model does not support the idea that friends forgo money in order to signal altruism. Rather (Proposition 2.5), if anything, choosing a price of zero is a signal of selfishness. This cuts against models like Bénabou and Tirole (2006) of forgoing money to seem prosocial. The divergent views arise because, in my model, friends are both buyers and sellers, rather than pure sellers being nice by sacrificing reward.

Fourth on the other hand, the model *does* support the idea of that friends forgo money to signal trust (Proposition 2.7). Friends—those who care about each other and trust each other's good motives—have a comparative advantage in helping each other without money, and this

comparative advantage gives rise to single-crossing that drives a separating equilibrium. In this particular model, this force is much weaker than the classical benefits of an efficient price, in that only when altruism and belief parameters are very extreme does it make the benefits of coordination worth the costs of weak incentives. Still, at least it demonstrates it is possible. Even so, the model does not support this interpretation very forcefully. Example 2.8 supports many other equilibria, including pooling (Example 2.12), as well as separating equilibria where the high type chooses only a slightly less preferred price (say, .49 instead of .50), and still signals her type.

Ultimately, while this model provides a game-theoretic explanation for why friends might avoid money, it also provides support for the classical result, that they really ought to pay each other. For the time being, friendship remains a puzzle for economists.

APPENDIX A

PROOFS

A.1 Proofs for Chapter 1 on the Marginal Value of Compensation

To assists in proofs of Propositions 1.1 and 1.3, I first establish the following Lemma about the best-response mappings:

Lemma A.1 (Bounded Best-Response)**.**

For any parameter values of $\alpha_b, \alpha_s, p \in [0, 01]$*, given best-response equations* (1.5)–(1.6) *(reprinted here)*

$$\kappa_b^* = (1 - \alpha_b)\, p + \alpha_b \, \frac{\int_0^{\kappa_s^*} v_s \, \mathrm{d}F_s(v_s)}{F_s(\kappa_s^*)} \tag{1.5}$$

$$\kappa_s^* = (1 - \alpha_s)\, p + \alpha_s \, \frac{\int_{\kappa_b^*}^{1} v_b \, \mathrm{d}F_b(v_b)}{1 - F_b\Big(\kappa_b^*\Big)}, \tag{1.6}$$

- *The buyer's best-response to a seller cutoff* κ_s *is in the interval* $[0, 1]$*, and is* < 1 *unless* $\alpha_b = 0$ *and* $p = 1$*.*

- *The seller's best-response to a buyer cutoff* κ_b *is in the interval* $[0, 1]$*, and is* > 0 *unless* $\alpha_s = 0$ *and* $p = 0$*.*

Proof of Lemma A.1 (Bounded Best-Response).

The proof proceeds by noting that each of these best-responses is a weighted average of two quantities, and then considering what values those two can take on.

To start, note (as pointed out in Remark 1.2) that the fraction $\frac{\int_0^{\kappa_s^*} v_s \, \mathrm{d}F_s(v_s)}{F_s(\kappa_s^*)}$ on the right-hand side of (1.5) is $\mathbb{E}\big[v_s \mid v_s \leq \kappa_s^*\big]$, the conditional expectation of v_s given that $v_s \leq$

κ_s^*, since its numerator $\int_0^{\kappa_s^*} v_s \, \mathrm{d}F_s(v_s) = \mathbb{E}\big[v_s \mid v_s \leq \kappa_s^*\big] \cdot \mathbb{P}\big[v_s \leq \kappa_s^*\big]$ and its denominator $F_s\big(\kappa_s^*\big) = \mathbb{P}\big[\kappa_s^*\big]$. Similarly, the fraction $\frac{\int_{\kappa_b^*}^1 v_b \, \mathrm{d}F_b(v_b)}{1-F_b\left(\kappa_b^*\right)}$ in (1.6) is equal to $\mathbb{E}\Big[v_b \mid v_b \geq \kappa_b^*\Big]$, the conditional expectation of the buyer's valuation v_b given that this valuation is above the buyer's cutoff κ_b^*.

Each of these conditional expectations must be within support $[0, 1]$, since they are each the average of valuations v_j in that support.

Moreover, the only way these can be at the boundary of their support is (in the limit) if the cutoff equals that boundary value. That is, $\mathbb{E}\big[v_s \mid v_s \leq \kappa_s^*\big] \equiv \mathbb{E}\big[v_s \mid 0 \leq v_s \leq \kappa_s^*\big]$ is > 0, unless $\kappa_s^* = 0$, in which case it $= 0$. And $\mathbb{E}\big[v_s \mid v_s \leq \kappa_s^*\big] < 1$, since the highest it could be is if $\kappa_s^* = 1$, but even then it would average realization $v_s = 1$ as well as strictly smaller realizations of v_s.

Similarly, $\mathbb{E}\Big[v_b \mid v_b \geq \kappa_b^*\Big] \equiv \mathbb{E}\Big[v_b \mid \kappa_b^* \leq v_b \leq 1\Big]$ is < 1, unless $\kappa_b^* = 1$, in which case it $= 1$. And $\mathbb{E}\Big[v_b \mid v_b \geq \kappa_b^*\Big] > 0$ for all κ_b^*.

Bearing these inequalities in mind, we see that in (1.5), κ_b^* is an average of p, with weight $(1 - \alpha_b)$, and of $\mathbb{E}\big[v_s \mid v_s \leq \kappa_s^*\big]$, with weight α_b. Since $p \in P = [0, 1]$ by assumption, this implies that the best-response κ_b^* is in $[0, 1]$, as the average of two quantities also in this interval. Further, since we have shown that $\mathbb{E}\big[v_s \mid v_s \leq \kappa_s^*\big] < 1$, the only way the best-response κ_b^* could be 1 would be if $p = 1$ and the weight on it were $(1 - \alpha_b) = 1$, in other words if $\alpha_b = 0$.

Analogously, in (1.6), the best-response κ_s^* is in $[0, 1]$, and it is > 0 unless $p = 0$ and the weight on it is $(1 - \alpha_s) = 1$, i.e. if $\alpha_s = 0$.

□

Proof of Proposition 1.1 (Non-Autarkic Equilibrium Cutoffs).

The best response by $i \in \{b, s\}$ to a cutoff κ_j, is a strategy where she agrees to trade when it is beneficial and declines to trade when it is not. This strategy calls for trading when the draw of v_i makes expected utility $U_i\Big(v_i, \kappa_j\Big)$ positive, and for declining to trade expected utility is negative. (When expected utility is zero, i is indifferent.)

Working through the equations for thms, the buyer, best-responding to some cutoff $\kappa_s > 0$, wants to trade if and only if

$$\begin{aligned} 0 \leq U_b(v_b, \kappa_s) &= \int_0^{\kappa_s} u_b(v_b, v_s)\, \mathrm{d}F_s(v_s) \\ &= \int_0^{\kappa_s} (v_b - \alpha_b v_s - (1-\alpha_b)p)\, \mathrm{d}F_s(v_s) \\ &= (v_b - (1-\alpha_b)\, p)F_s(\kappa_s) - \alpha_b \int_0^{\kappa_s} v_s\, \mathrm{d}F_s(v_s) \\ \Leftrightarrow v_b \geq (1-\alpha_b)\, p + \alpha_b &\frac{\int_0^{\kappa_s} v_s\, \mathrm{d}F_s(v_s)}{F_s(\kappa_s)}. \end{aligned} \tag{A.1}$$

The right-hand side of (A.1) is exactly κ_b^* from (A.1), establishing that this is indeed the best-response. By Lemma A.1, this cutoff is in $[0, 1]$

Analogously, s prefers to trade when

$$\begin{aligned} 0 \leq U_s(v_s, \kappa_b) &= \int_{\kappa_b}^1 u_s(v_s, v_b)\, \mathrm{d}F_b(v_b) \\ \Leftrightarrow v_s \leq (1-\alpha_s)\, p + \alpha_s &\frac{\int_{\kappa_b}^1 v_b\, \mathrm{d}F_b(v_b)}{1 - F_b(\kappa_b)}. \end{aligned} \tag{A.2}$$

The right-hand side of (A.2) is the optimal cutoff, proving that (1.6) represents a best-response by s to κ_b^*. Thus, the two equations are mutual best-responses, and hence characterize an equilibrium.

□

Proof of Proposition 1.3 (Equilibrium Existence and Uniqueness).

First, we show an autarkic equilibrium always exists. If s never agrees to trade after any valuation, then b's interim utility is 0 no matter what strategy she plays, so refusing to trade is among her best responses. Similarly, if b never agrees to trade, then it is a best response for s to refuse to trade as well. Hence, for any parameters, there exists an autarkic equilibrium. Note that this equilibrium does not satisfy (1.5)–(1.6), since those equations already assume a positive probability of trade (in order for their denominators to be nonzero).

Next, we show that an equilibrium exists satisfying (1.5)–(1.6), that it is non-autarkic, and that it is unique under (1.9).

Consider the twice-iterated best-response by b, defined by starting with a buyer cutoff κ_b, obtaining the seller's best-response to it via (1.8), and then obtaining buyer's best response to *that* via (1.7):

$$\begin{aligned}\varphi(\kappa_b) &:= \kappa_b^*\big(\kappa_s^*(\kappa_b)\big) \\ &= (1-\alpha_b)p + \alpha_b \mathbb{E}\left[v_s \mid v_s \leq \Big((1-\alpha_s)p + \alpha_s \mathbb{E}\big[v_b \mid v_b \geq \kappa_b\big]\Big)\right]. \qquad \text{(A.3)}\end{aligned}$$

Any fixed point κ_b (satisfying $\kappa_b = \varphi(\kappa_b)$), along with the best-response seller cutoff $\kappa_s^*(\kappa_b)$, would constitute an equilibrium to the game, since they would be mutual best-responses. Thus, to prove the proposition, we show that a fixed point exists under the specified conditions, and that it is unique under condition (1.9).

Equilibrium existence is proved using Brouwer's fixed-point theorem. By Lemma A.1, each of the best-response map $[0,1]$ to $[0,1]$. Moreover, they are trivially continuous, because all of the functions with them are continuous. So, composing these functions also keeps the output within $[0,1]$. Hence $\varphi(\cdot)$ is a continuous mapping from $[0,1]$ to $[0,1]$, and so has a fixed point by Brouwer's theorem. This establishes that equilibrium exists.

Next, to show that equilibrium is non-autarkic under the conditions specified, note that autarky involves $(1 - F_b(\kappa_b)) \cdot F_s(\kappa_s)$, and so requires either $\kappa_b = 1$ or $\kappa_s = 0$. Based on Lemma A.1, if $p \in (0, 1)$ strictly, then the buyer and the seller both will play an interior cutoff. So, a necessary condition for autarky is $p \in \{0, 1\}$. If a player's altruism is 0, then her best-response is simply to set her cutoff equal to the price. Thus, if $p = 0$, autarky (with $\kappa_s^* = 0$) can occur if $\alpha_s = 0$, and similarly, if $p = 1$, autarky (with $\kappa_b^* = 1$) can occur of $\alpha_b = 0$. Otherwise, they will trade with positive probability

To show uniqueness, notice that the LHS expression in condition (1.9) is the derivative of the iterated best-response function (A.3) with respect to κ_b, straightforwardly from applying the chain rule.[1] The inequality (1.9) ensures that the twice-iterated best-response function has a slope < 1, and so intersects the 45° line just once. Hence, there is a unique fixed point, i.e. a unique equilibrium.

□

Proof of Lemma 1.4 (Cutoff Strategies).

Suppose j plays an arbitrary strategy (a decision rule telling the probability she agrees to trade following each possible draw of valuation v_j), which is not necessarily a cutoff strategy. Then, expected utiltiy (1.1)–(1.2) for a player i (of either role b or s) given valuation draw v_i is

$$\mathbb{E}\Big[u_i\Big(v_i, v_j\Big) \mid j \text{ agrees to trade}\Big] \cdot \mathbb{P}[j \text{ agrees to trade}]\,.$$

In non-autarkic equilibrium, j agrees to trade with positive probability, so this expression is nonzero for at least some v_i. So to maximize it, i's best-response is to trade at valuations v_i where it is positive and not trade where it is negative. (Also, as a convention we dictate that

1. The derivative $\frac{d}{dx}\mathbb{E}_{v_b \sim F_b}\big[v_b \mid v_b \geq x\big]$ is expanded in equation (A.16), in the proof of Proposition 1.8.

she agree to trade if it is zero.) So, expanding it out, and using the fact that the expectations operator $\mathbb{E}\big[\cdot \mid \text{agrees to trade}\big]$ and the utility functions $u_i(\cdot)$ commute, the generalization of best-response is

$$\begin{aligned} b \text{ agrees to trade iff} \quad & v_b \geq (1-\alpha_b)\, p + \alpha_b\, \mathbb{E}\big[v_s \mid s \text{ agrees to trade}\big] \\ s \text{ agrees to trade iff} \quad & v_s \leq (1-\alpha_s)\, p + \alpha_s\, \mathbb{E}\big[v_b \mid b \text{ agrees to trade}\big]\,. \end{aligned}$$

These expressions mean that best-responses to a partner who trades with positive probability must be a cutoff strategy, with cutoff given by the RHS. Hence, in any non-autarkic equilibrium, equilibrium strategies must be cutoffs, as was to be shown. □

Proof of Proposition 1.6 (Comparative Statics on p).
In first order conditions (1.7)–(1.8), the partial derivative of cutoff strategy with respect to price is $\frac{\partial \kappa_i^*}{\partial p} = 1 - \alpha_i$. Since a higher p increases both of these best-response curves, and since by Proposition 1.5 they are both upward-sloping curves, then their new intersection is at an equilibrium with higher cutoffs. □

Proof of Proposition 1.7 (Comparative Statics on Altruism).
At $p = 0$, the first-order conditions (1.7)–(1.7) imply that the partial derivatives with respect to own and partner's altruism are

$$\begin{aligned} \frac{\partial \kappa_b^*}{\partial \alpha_b} &= \mathbb{E}\big[v_s \mid v_s \leq \kappa_s^*\big] & \frac{\partial \kappa_b^*}{\partial \alpha_s} &= 0 \\ \frac{\partial \kappa_s^*}{\partial \alpha_s} &= \mathbb{E}\big[v_b \mid v_b \geq \kappa_b^*\big] & \frac{\partial \kappa_s^*}{\partial \alpha_b} &= 0. \end{aligned}$$

The nonzero expressions are nonnegative because the valuation supports are assumed to be nonnegative. Then, as in Proposition 1.6's proof, an increase in these best-response curves increases their intersection, because by Proposition 1.5 they are both upward-sloping curves.

Since their intersection defines the equilibrium, this implies that equilibrium cutoffs are increasing in altruism as well.

□

Proof of Proposition 1.8 (Marginal Value of Compensation to Seller and Buyer).

Showing $W_s'(0) \geq 0$ for all $\alpha_s, \alpha_b \in [0,1]$. We begin from the definition of $W_s(p)$ (1.17) and totally differentiate, decomposing into three terms:

$$\begin{aligned} W_s(p) &= \mathcal{U}_s\big(\kappa_s^*(p)\,,\,\kappa_b^*(p)\,;p\big) \\ \implies \frac{d}{dp}W_s(p) &= \underbrace{\frac{\partial}{\partial\kappa_s}\mathcal{U}_s \cdot \frac{d}{dp}\kappa_s^*(p)}_{=0 \text{ by Envelope}} + \underbrace{\frac{\partial}{\partial\kappa_b}\mathcal{U}_s \cdot \frac{d}{dp}\kappa_b^*(p)}_{\text{Marginal}} + \underbrace{\frac{\partial}{\partial p}\mathcal{U}_s}_{\text{Inframarginal}} \quad . \end{aligned} \tag{A.4}$$

The first term is zero by the Envelope Theorem, since s's first order condition demands that $\frac{\partial}{\partial\kappa_s}\mathcal{U}_s = 0$. The second term represents marginal b valuations who become willing or unwilling to trade, as b's cutoff κ_b changes to reflect the p. The third term represents change to s's value from a higher price to inframarginal valuations v_s and v_b. We now show that the second and third terms are nonnegative.

The inframarginal term is simple to compute. First recall that the effect of changing price on ex-post utility is independent of valuations: $\frac{\partial}{\partial p}u_s(v_s, v_b; p) = \frac{\partial}{\partial p}(-v_s + \alpha_s v_b + (1-\alpha_s)p) = 1 - \alpha_s$. Then, by the formula $\mathcal{U}_s(\kappa_s, \kappa_b) = \int_0^{\kappa_s}\int_{\kappa_b}^1 u_s(v_s, v_b; p)\,\mathrm{d}F_b(v_b)\,\mathrm{d}F_s(v_s)$ for ex-ante

utility (1.4), changing p without changing cutoffs accomplishes

$$\begin{aligned}\frac{\partial}{\partial p}\mathcal{U}_s(\kappa_s,\kappa_b;p) &= \int_0^{\kappa_s}\int_{\kappa_b}^1 \frac{\partial}{\partial p}u_s(v_s,v_b;p)\,\mathrm{d}F_b(v_b)\,\mathrm{d}F_s(v_s)\\ &= \int_0^{\kappa_s}\int_{\kappa_b}^1 (1-\alpha_s)\,\mathrm{d}F_b(v_b)\,\mathrm{d}F_s(v_s)\\ &= (1-F_s(\kappa_s))\cdot F_b(\kappa_b)\cdot(1-\alpha_s).\end{aligned}$$

This expression equals $(1-\alpha)$, the impact of changing price on ex-post utility, times the probability of trade $(1-F_s(\kappa_s))\cdot F_b(\kappa_b)$. It is nonnegative, and is strictly positive whenever $\alpha_s<1$ and $\mathbb{P}[\text{trade}]>0$.

To attack the marginal term, we first note that $\frac{d}{dp}\kappa_b^*(p)\geq 0$ by Proposition 1.6. Then, we relate s's ex-ante utility $\mathcal{U}_s$ to b's ex-ante utility $\mathcal{U}_b$, and piggyback off b's first order condition. Observe that the two players' ex-ante utility expressions are nearly identical, only with different weights on their ex-ante material payoffs:

$$\mathcal{U}_s(\kappa_s,\kappa_b;p) = \ 1\cdot\int_0^{\kappa_s}\int_{\kappa_b}^1(-v_s+p)\,\mathrm{d}F_b(v_b)\,\mathrm{d}F_s(v_s)+\alpha_s\cdot\int_0^{\kappa_s}\int_{\kappa_b}^1(\quad v_b-p)\,\mathrm{d}F_b(v_b)\,\mathrm{d}F_s(v_s)$$

$$\mathcal{U}_b(\kappa_b,\kappa_s;p) = \alpha_b\cdot\int_0^{\kappa_s}\int_{\kappa_b}^1(-v_s+p)\,\mathrm{d}F_b(v_b)\,\mathrm{d}F_s(v_s)+\ 1\cdot\int_0^{\kappa_s}\int_{\kappa_b}^1(\quad v_b-p)\,\mathrm{d}F_b(v_b)\,\mathrm{d}F_s(v_s)\,.$$

The buyer's first-order condition $\frac{\partial}{\partial\kappa_b}\mathcal{U}_b=0$ implies that

$$1\cdot\frac{\partial}{\partial\kappa_b}\int_0^{\kappa_s}\int_{\kappa_b}^1(\quad v_b-p)\,\mathrm{d}F_b(v_b)\,\mathrm{d}F_s(v_s) = -\alpha_b\cdot\frac{\partial}{\partial\kappa_b}\int_0^{\kappa_s}\int_{\kappa_b}^1(-v_s+p)\,\mathrm{d}F_b(v_b)\,\mathrm{d}F_s(v_s)\,,$$

and then substituting this identity in to $\frac{\partial}{\partial \kappa_b}\mathcal{U}_s$ and employing Leibniz's rule gets us

$$\begin{aligned}
\frac{\partial}{\partial \kappa_b}\mathcal{U}_s = \quad & 1 \cdot \qquad \frac{\partial}{\partial \kappa_b}\int_0^{\kappa_s}\int_{\kappa_b}^{1}(-v_s+p)\,\mathrm{d}F_b(v_b)\,\mathrm{d}F_s(v_s) \\
& + \alpha_s\Big(- \alpha_b \cdot \frac{\partial}{\partial \kappa_b}\int_0^{\kappa_s}\int_{\kappa_b}^{1}(-v_s+p)\,\mathrm{d}F_b(v_b)\,\mathrm{d}F_s(v_s)\Big) \\
= \ & (1-\alpha_s\alpha_b)\cdot\frac{\partial}{\partial \kappa_b}\int_0^{\kappa_s}\int_{\kappa_b}^{1}(-v_s+p)\,\mathrm{d}F_b(v_b)\,\mathrm{d}F_s(v_s) \\
= \ & (1-\alpha_s\alpha_b)\cdot -f_b(\kappa_b)\cdot\int_0^{\kappa_s}(-v_s+p)\,\mathrm{d}F_b(v_b)\,.
\end{aligned}$$

At $p=0$, this is nonnegative for two reasons. First, $(1-\alpha_s\alpha_b)\geq 0$ because $\alpha_s,\alpha_b\leq 1$. Second, at $p=0$, $-\int_0^{\kappa_s}(-v_s+0)\,\mathrm{d}F_b(v_b) = \mathbb{E}\big[v_s \mid v_s\leq\kappa_s\big]\cdot\mathbb{P}[v_s\leq\kappa_s]$, which is nonnegative because all supported v_s are nonnegative, and is strictly positive so long as s is willing to trade with positive probability.

Putting these three terms together, we see that they are all nonnegative (and are strictly positive except under autarky or total selflessness), making the overall marginal value of compensation nonnegative to the seller. This proves that $W_s'(0)\geq 0$.

Showing $W_b'(0)>0$ for α_s close to 0, and $W_b'(0)<0$ for $\alpha_s=1,\alpha_b=0$. We prove these two results together, since many of the steps are the same. To do so, we will totally differentiate $\mathcal{U}_b$ and analyze the components, and then make use of the equilibrium strategies at these particular altruism parameter values.

Start with solving the equilibrium in these two cases, using best-response equations (1.7)–(1.8). At $\alpha_s=0$, and with p respectively general (left equations) and $=0$ (right equations),

equilibrium cutoffs are

$$\kappa_s^*(p;0,\alpha_b) = \underbrace{(1-\alpha_s)}_{=1}\cdot p + \underbrace{\alpha_s}_{=0}\cdot \mathbb{E}\big[v_b \mid v_b \geq \kappa_b\big] = p \qquad \kappa_s^*(0;0,\alpha_b) = 0 \tag{A.5}$$

$$\kappa_b^*(p;\alpha_b,0) = (1-\alpha_b)\cdot p + \alpha_b \cdot \mathbb{E}\big[v_s \mid v_s \leq \underbrace{\kappa_s}_{=p}\big] \qquad \kappa_b^*(0;\alpha_b,0) = 0. \tag{A.6}$$

In other words, at $p = 0$, there is autarky because the seller sells with zero probability (she sells only if $v_s = 0$).

Next, at $\alpha_b = 0, \alpha_s = 1$, and p either general (left equations) or $= 0$ (right equations), equilibrium cutoffs are

$$\kappa_s^*(p;1,0) = \underbrace{(1-\alpha_s)}_{=0}\cdot p + \underbrace{\alpha_s}_{=1}\cdot \mathbb{E}\big[v_b \mid v_b \geq \underbrace{\kappa_b}_{=p}\big] \qquad \kappa_s^*(0;1,0) = \mathbb{E}[v_b] \tag{A.7}$$

$$\kappa_b^*(p;0,1) = \underbrace{(1-\alpha_b)}_{=1}\cdot p + \underbrace{\alpha_b}_{=0}\cdot \mathbb{E}\big[v_s \mid v_s \leq \kappa_s\big] = p \qquad \kappa_b^*(0;0,1) = 0. \tag{A.8}$$

Note that in $\kappa_s^*(0;0,\alpha_b)$, we have simplified $\mathbb{E}\big[v_b \mid v_b \geq 0\big] = \mathbb{E}[v_b]$, because v_b is always ≥ 0 by assumption.

With these equilibrium solutions in hand, we turn to the decomposition of the marginal value of compensation:

$$\begin{aligned} W_s(p) &= \mathcal{U}_b\big(\kappa_b^*(p)\,,\,\kappa_s^*(p)\,;p\big) \\ \implies \frac{d}{dp}W_b(p) &= \underbrace{\frac{\partial}{\partial\kappa_b}\mathcal{U}_b \cdot \frac{d}{dp}\kappa_b^*(p)}_{=0 \text{ by Envelope}} + \underbrace{\frac{\partial}{\partial\kappa_s}\mathcal{U}_b \cdot \frac{d}{dp}\kappa_s^*(p)}_{\text{Marginal}} + \underbrace{\frac{\partial}{\partial p}\mathcal{U}_b}_{\text{Inframarginal}} \quad . \end{aligned} \tag{A.9}$$

The first term here is zero—for any parameter values—because $\frac{\partial}{\partial\kappa_b}\mathcal{U}_b = 0$ by b's first order condition (just as was $\frac{\partial}{\partial\kappa_s}\mathcal{U}_s$ above in (A.4)).

The third (inframarginal) term is similar to the inframarginal term in the analysis of $W_s'(0)$ above. Specifically, the effect of price on ex-post utility is again independent of valuation: $\frac{\partial}{\partial p}u_b(v_b, v_s; p) = \frac{\partial}{\partial p}(v_b + \alpha_b v_s - (1-\alpha_b)p) = -(1-\alpha_b)$. Since this effect changes utility whenever trade takes place,

$$\frac{\partial}{\partial p}\mathcal{U}_b(\kappa_b, \kappa_s; p) = -(1-\alpha_b) \cdot (1 - F_b(\kappa_b)) \cdot F_s(\kappa_s)\,. \tag{A.10}$$

This expression is $(1-\alpha_b) \cdot \mathbb{P}[\text{trade}]$. It is weakly negative, and strictly negative except under autarky. For parameter values $p = 0, \alpha_s = 0$, this term (A.10) works out to

$$\left.\frac{\partial}{\partial p}\right|_{p=0} \mathcal{U}_b(0, 0; p) = 0 \tag{A.11}$$

by (A.5)–(A.6), since the seller never sells ($\kappa_s^* = 0$).

But for parameter values $p = 0, \alpha_s = 1, \alpha_b = 0$, the expression (A.10) is

$$\left.\frac{\partial}{\partial p}\right|_{p=0} \mathcal{U}_b\big(0, \mathbb{E}[v_b]\,; p\big) = -F_s\big(\mathbb{E}[v_b]\big) \tag{A.12}$$

since $1 - \alpha_b = 1$, and by (A.7)–(A.8) the buyer always buys ($\kappa_b^* = 0$).

Now it's time for the second (marginal) term in $\frac{d}{dp}W_b(p)$ (A.9). It is the product $\frac{\partial}{\partial \kappa_s}\mathcal{U}_b \cdot \frac{d}{dp}\kappa_s^*(p)$. We expand the first factor with Leibniz's rule:

$$\begin{aligned}\frac{\partial}{\partial \kappa_s}\mathcal{U}_b(\kappa_b, \kappa_s; p) &= \frac{\partial}{\partial \kappa_s}\int_0^{\kappa_s}\int_{\kappa_b}^1 (v_b - p + \alpha_b \cdot (-v_s + p))\,\mathrm{d}F_b(v_b)\,\mathrm{d}F_s(v_s) \\ &= f_s(\kappa_s) \cdot \int_{\kappa_b}^1 (v_b - p + \alpha_b \cdot (-\kappa_s + p))\,\mathrm{d}F_b(v_b)\,.\end{aligned}$$

In the case of $p = 0, \alpha_s = 0$, we plug in equilibrium solutions (A.5)–(A.6) to arrive at

$$\begin{aligned}\left.\frac{\partial}{\partial \kappa_s}\right|_{\kappa_s=0} \mathcal{U}_b(0, \kappa_s; 0) &= f_s(0) \cdot \int_0^1 (v_b - 0 + \alpha_b \cdot (-0 + 0))\, \mathrm{d}F_b(v_b) \\ &= f_s(0)\, \mathbb{E}[v_b]\,. \end{aligned} \tag{A.13}$$

In the case of $p = 0, \alpha_s = 1, \alpha_b = 0$, we plug in equilibrium solutions (A.7)–(A.8) to arrive at

$$\begin{aligned}\left.\frac{\partial}{\partial \kappa_s}\right|_{\kappa_s=\mathbb{E}[v_b]} \mathcal{U}_b(0, \kappa_s; 0) &= f_s\big(\mathbb{E}[v_b]\big) \cdot \int_0^1 (v_b - 0 + 0 \cdot (-\mathbb{E}[v_b] + 0))\, \mathrm{d}F_b(v_b) \\ &= f_s\big(\mathbb{E}[v_b]\big) \cdot \mathbb{E}[v_b] \cdot \underbrace{(1 - F_b(0))}_{=1} \end{aligned} \tag{A.14}$$

We can directly compute the second half of the inframarginal term, the expression $\frac{d}{dp}\kappa_s^*(p)$. By (A.5), when $\alpha_s = 0$, $\kappa_s^*(p; \alpha_s, \alpha_b) = p$, so

$$\frac{d}{dp}\kappa_s^*(p; 0, \alpha_b) = 1. \tag{A.15}$$

And when $\alpha_s = 1, \alpha_b = 0$, by (A.7), $\kappa_s^*(p; \alpha_s; \alpha_b) = \mathbb{E}\big[v_b \mid v_b \geq p\big]$. We compute its derivative directly:

$$\begin{aligned} \mathbb{E}\big[v_b \mid v_b \geq p\big] &= \frac{\int_p^1 v_b\, \mathrm{d}F_b(v_b)}{1 - F_b(p)} \\ \implies \frac{d}{dp}\mathbb{E}\big[v_b \mid v_b \geq p\big] &= \frac{(-f_b(p)\, p) \cdot (1 - F_b(p)) - \int_p^1 v_b\, \mathrm{d}F_b(v_b) \cdot (-f_b(p))}{(1 - F_b(p))^2} \\ &= \frac{f_b(p)}{1 - F_b(p)} \cdot (-p + \mathbb{E}\big[v_b \mid v_b \geq p\big]) \end{aligned} \tag{A.16}$$

$$\begin{aligned} \implies \left.\frac{d}{dp}\right|_{p=0} \kappa_s^*(p; 1, 0) &= \left.\frac{d}{dp}\right|_{p=0} \mathbb{E}\big[v_b \mid v_b \geq p\big] \\ &= \frac{f_b(0)}{1 - F_b(0)} \cdot (\mathbb{E}[v_b]) \end{aligned} \tag{A.17}$$

Now, having expanded out all the terms, we combine them back to form decomposition (A.9) of $W_b'(0)$. At parameter values $\alpha_s = 0, p = 0$, the first term is zero by envelope, the second inframarginal term is $\frac{\partial}{\partial \kappa_s}\mathcal{U}_b \cdot \frac{d}{dp}\kappa_s^*(p) = (f_s(0) \cdot \mathbb{E}[v_b]) \cdot 1$ by (A.13) and (A.15), and the third inframarginal term is 0 by (A.11). That is,

$$W_b'(0; \alpha_b, 0) = f_s(0) \cdot \mathbb{E}[v_b] \geq 0,$$

with > 0 under the assumption that $f_s(0) > 0$, as was to be shown at $\alpha_s = 0$. If this is strictly positive at $\alpha_s = 0$, then it is also positive for values sufficiently close to 0, since all the equations are continuous.

At parameter values $\alpha_s = 1, \alpha_b = 0, p = 0$, we combine the marginal term from (A.14) and (A.17) with the inframarginal term from (A.12), and we do some manipulations to get a sufficient condition for a negative marginal value of compensation:

$$\begin{aligned}
\frac{d}{dp}W_b(p) &= \underbrace{\frac{\partial}{\partial \kappa_b}\mathcal{U}_b \cdot \frac{d}{dp}\kappa_b^*(p)}_{=0 \text{ by Envelope}} + \underbrace{\frac{\partial}{\partial \kappa_s}\mathcal{U}_b \cdot \frac{d}{dp}\kappa_s^*(p)}_{\text{Marginal}} + \underbrace{\frac{\partial}{\partial p}\mathcal{U}_b}_{\text{Inframarginal}} \quad . \\
W_b'(0; 0, 1) &= 0 + \Big(f_s\big(\mathbb{E}[v_b]\big) \cdot \mathbb{E}[v_b]\Big) \cdot \left(\frac{f_b(0)}{1 - F_b(0)} \cdot (\mathbb{E}[v_b])\right) - F_s\big(\mathbb{E}[v_b]\big)\,, \\
\text{so, } W_b'(0; 0, 1) < 0 \quad &\text{iff} \quad 1 > \frac{f_s(v_b)}{F_s\big(\mathbb{E}[v_b]\big)} \cdot \frac{f_b(0)}{1 - F_b(0)} \cdot \mathbb{E}[v_b]^2\,.
\end{aligned}$$

This is condition (1.18), as was to be shown for parameter values $\alpha_s = 1, \alpha_b = 0$. □

Proof of Proposition 1.9 (Marginal Value of Compensation under Uniform Valuations).

Define the buyer's altruism threshold function $\zeta_b(\cdot)$ as

$$\zeta_b(\alpha_s) := -\frac{3}{2} + \frac{1}{\alpha_s} + \frac{1}{2}\sqrt{3}\sqrt{3 + \frac{4}{\alpha_s} - \frac{4}{\alpha_s^2}}. \tag{A.18}$$

We show that the sign of $W_b'\Big(0; \alpha_b, \alpha_j\Big)$ is determined by $\alpha_b \gtrless \zeta_b(\alpha_s)$, by solving the inequality analytically.

We begin by computing the value of the partnership to b. We plug in (1.14)–(1.15) for equilibrium cutoffs under uniformly distributed valuations into (1.16) for $W_b(p)$, and differentiate with respect to p at $p = 0$. The result is

$$\begin{aligned} W_b(p) &= \mathcal{U}_b\big(\kappa_b^*(p)\,,\,\kappa_s^*(p)\,;p\big) \\ &= \frac{4(\alpha_s(p\alpha_b + p - 1) - 2p)(\alpha_b((p-1)\alpha_s + p) - 2p + 2)^2}{(\alpha_b\alpha_s - 4)^3} \\ W_b'(0) &= \frac{4(\alpha_b\alpha_s - 2)\left(\alpha_b(\alpha_b+3)\alpha_s^2 - 2(\alpha_b+3)\alpha_s + 4\right)}{(\alpha_b\alpha_s - 4)^3}. \end{aligned}$$

Ignoring the denominator, this expression is cubic in α_b, so solving for the value of α_b that makes $W_b'(0) = 0$ yields three roots:

$$\alpha_b = \frac{2}{\alpha_s} \quad \text{and} \quad \alpha_b = -\frac{3}{2} + \frac{1}{\alpha_s} \pm \frac{1}{2}\sqrt{3}\sqrt{3 + \frac{4}{\alpha_s} - \frac{4}{\alpha_s^2}}.$$

To sign $W_b'(0)$, note that at $\alpha_b = 0$, $W_b'(0) = \frac{\frac{2}{3} - \alpha_s}{12}$, which is positive iff $\alpha_s < \frac{2}{3}$. Then as α_b rises, the sign flips whenever α_b crosses one of the above roots. The first root $\frac{2}{\alpha_s}$ is > 1 and so is out of domain $[0, 1]$, while the second two are real only when $\alpha_s \geq \frac{2}{3}$. So, the overall expression $W_b'(0)$ is positive for $\alpha_s < \frac{2}{3}$. Of the "$\pm$" pair of roots, the "$-$" root is negative for $\alpha_s \in [\frac{2}{3}, 1]$, but the "$+$" root is in $[0, 1]$. When α_b is in between these two roots, $W_b'(0) > 0$. We therefore use the "$+$" root as the definition of the altruism threshold function $\zeta_b(\alpha_s)$. Within interval $[\frac{2}{3}, 1]$ for α_b, the condition for $W'(0) > 0$ is $\alpha_b < \zeta_b(\alpha_s)$.

□

Proof of Proposition 1.10 (Marginal Value of Compensation with Random Roles).
These proofs employ and build off of Proposition 1.8. To sign W', we sign the two fixed-role marginal values of compensation W'_b and W'_s, of which it is an average.

Proof of $\alpha_j \approx 0, \alpha_i < 1$ case. We show that $W'\Big(0; \alpha_i, \alpha_j\Big) > 0$ in this case.

First, when i is the seller, $W'_s\Big(0; \alpha_i, \alpha_j\Big) \geq 0$, and second, when i is the buyer, $W'_b\Big(0; \alpha_i, \alpha_j\Big) \geq 0$ if α_j is sufficiently close to 0, both by Proposition 1.8, Thus, the average of these two is nonnegative as well, for α_j sufficiently close to 0.

Proof of $\alpha_j = 1, \alpha_i = 0$ case. We show that under these parameter values, the condition for $W'\Big(0; \alpha_i, \alpha_j\Big)$ is inequality (1.20). Using $W'\Big(0; \alpha_i, \alpha_j\Big) = \frac{1}{2}W'_b\Big(0; \alpha_i, \alpha_j\Big) + \frac{1}{2}W'_s\Big(0; \alpha_i, \alpha_j\Big)$, we show that at these parameter values, $W'_s\Big(0; \alpha_i, \alpha_j\Big) = 0$, so that $W'(0; 0, 1) < 0$ iff $W'_b(0; 0, 1) < 0$; this way, condition (1.20) for negative random-role marginal value of compensation is identical to condition (1.18) for negative buyer's marginal value of compensation from Proposition 1.8.

To show $W'_s\Big(0; \alpha_i, \alpha_j\Big) = 0$ at $\alpha_j = 1, \alpha_i = 0$, we start with the equilibrium cutoffs. Plugging in to the above-computed cutoffs (A.5)–(A.6), these are $\kappa^*_s(p; 0, 1) = p$ and $\kappa^*_b(p; 1, 0) = \mathbb{E}\big[v_s \mid v_s \leq p\big]$. At $p = 0$, these cutoffs mean autarky, because the selfish seller i refuses to trade since she is not compensated for her trouble.

Next, we totally differentiate $W_s(p)$, and split into three terms, as in (A.4). The first term is zero since $\frac{\partial \mathcal{U}_s}{\partial \kappa_s} = 0$ by the seller's first order condition. The third (inframarginal) term is zero because it is $(1 - \alpha_i)\mathbb{P}[\text{trade}]$, with probability of trade being zero here.

For the second (marginal) term, we take differentiate with Leibniz's rule and then plug in

the equilibrium cutoffs.

$$\begin{aligned}\left.\frac{\partial}{\partial \kappa_b}\right|_{\kappa_b=\kappa_b^*} W_s\big(p;\kappa_s^*,\kappa_b\big) &= \left.\frac{\partial}{\partial \kappa_b}\right|_{\kappa_b=\kappa_b^*} \int_{\kappa_b}^{1}\int_{0}^{\kappa_s^*} (-v_s + p + \alpha_i(v_b - p))\,\mathrm{d}F_s(v_s)\,\mathrm{d}F_b(v_b) \\ &= -f_b(\kappa_b)\int_0^{\kappa_s^*} (-v_s + p + \alpha_i(v_b - p))\,\mathrm{d}F_s(v_s) \\ &= 0\end{aligned}$$

because seller i's equilibrium cutoff is $\kappa_s^* = 0$, making the inner integral 0, since its bounds of integration are from 0 to $\kappa_s^* = 0$. Because i never sells, she does not gain on the margin if j becomes a bit more willing to buy.

Since all three components of i's seller marginal value of compensation are zero, overall we have $W_s'(0) = 0$.

Therefore, with i totally selfish and j totally altruistic, $W'\Big(0;\alpha_i,\alpha_j\Big) = \frac{1}{2}W_b\Big(0;\alpha_i,\alpha_j\Big) + 0$. So they have the same sign, and in particular the have the same condition for being negative.

□

Proof of Proposition 1.11 (Marginal Value of Compensation with Random Roles under Uniform Valuations).
Define the random-role altruism threshold function $\zeta(\cdot)$ as

$$\zeta\Big(\alpha_j\Big) := \frac{-3\alpha_j^2 + 5\alpha_j - 6 + \sqrt{3}\sqrt{3\alpha_j^4 - 2\alpha_j^3 - 9\alpha_j^2 - 4\alpha_j + 12}}{2\alpha_j^2 - 6\alpha_j}. \tag{A.19}$$

We show that the sign of $W'\Big(0;\alpha_i,\alpha_j\Big)$ is determined by $\alpha_i \gtrless \zeta\Big(\alpha_j\Big)$, by solving the inequality analytically.

We begin by computing the value of the partnership to i: starting from formula (1.19) for W in terms of W_b and W_s, we then expand with equations (1.16)–(1.17) for W_b and W_s and

(1.10)–(1.11) for $\mathcal{U}_b$ and $\mathcal{U}_s$, and lastly plug in closed-form uniform equilibrium expressions (1.14)–(1.15). This value, and the marginal value of compensation (its derivative evaluated at $p = 0$) are

$$\begin{aligned} W\Big(p;\alpha_i,\alpha_j\Big) = \quad & \frac{2(\alpha_j(p\alpha_i+p-1)-2p)(\alpha_i((p-1)\alpha_j+p)-2p+2)^2}{(\alpha_i\alpha_j-4)^3} \\ & - \frac{2(\alpha_j((p-1)\alpha_i+p)-2p+2)(\alpha_i(p\alpha_j+p-1)-2p)^2}{(\alpha_i\alpha_j-4)^3} \end{aligned} \tag{A.20}$$

$$\begin{aligned} W'\Big(0;\alpha_i,\alpha_j\Big) = & \frac{1}{(\alpha_i\alpha_j-4)^3} \cdot 2(\alpha_i\alpha_j+\alpha_i-2)\cdot \\ & \Big(\alpha_i^2\alpha_j^2-3\alpha_i^2\alpha_j+3\alpha_i\alpha_j^2-5\alpha_i\alpha_j+6\alpha_i-6\alpha_j+4\Big). \end{aligned} \tag{A.21}$$

To sign $W'(0)$, note that the denominator and the first upstairs factor are negative for $\alpha_i, \alpha_j \in [0,1]$. So for the overall $W'(0)$ to be positive, the second factor must be positive. This second factor is quadratic in α_i, and (since $\alpha_j < 1$) it is positive for α_i between its roots

$$\frac{-3\alpha_j^2+5\alpha_j-6 \pm \sqrt{3}\sqrt{3\alpha_j^4-2\alpha_j^3-9\alpha_j^2-4\alpha_j+12}}{2\alpha_j^2-6\alpha_j}.$$

This means that $W'(0) > 0$ so long as α_i is in between those two roots. For $\alpha_j < \frac{2}{3}$, these two roots straddle $[0,1]$, so any $\alpha_j \in [0,1]$ is in between the roots, making $W'(0) > 0$. For $\alpha_i \in \left[\frac{2}{3},1\right]$, the "$-$" root—which is what $\zeta\Big(\alpha_j\Big)$ is defined to be (A.19)—is in $[0,1]$, but the "$+$" root is still > 1. So, for $\alpha_i \geq \frac{2}{3}$, the condition for $W'(0) > 0$ is that $\alpha_i > \zeta\Big(\alpha_j\Big)$. That is: $W'(0) > 0$ if $\alpha_i > \zeta\Big(\alpha_j\Big)$, $W'(0) < 0$ if $\alpha_i < \zeta\Big(\alpha_j\Big)$, and $W'(0) = 0$ if $\alpha_i = \zeta\Big(\alpha_j\Big)$, which of course was to be shown.

It is a straightforward computation that $\zeta\Big(\frac{2}{3}\Big) = 0$ and $\zeta\Big(1\Big) = 1$. It is less straightforward,

but still a workable computation, to show that $\zeta'\Big(\alpha_j\Big) > 0$. This derivative is

$$\begin{aligned}\zeta'\Big(\alpha_j\Big) = &\frac{1}{(\alpha_j-3)^2\alpha_j^2\sqrt{3\alpha_j^4-2\alpha_j^3-9\alpha_j^2-4\alpha_j+12}}\\ &\cdot\Bigg(\sqrt{3}\left(-4\alpha_j^4+6\alpha_j^3+3\alpha_j^2-15\alpha_j+18\right)\\ &\quad+\left(2\alpha_j^2+6\alpha_j-9\right)\sqrt{3\alpha_j^4-2\alpha_j^3-9\alpha_j^2-4\alpha_j+12}\Bigg),\end{aligned}$$

and after some wrangling it can be shown to be positive for $\alpha_j \in \left[\frac{2}{3}, 1\right]$. This is because the numerator sums a positive and a negative term, but overall it is positive if α_j is in the intervals between real roots of two polynomials. These intervals are approximately $[.650, .136]$ and $[-.1444, 1.713]$, so α_j is inside them because, by assumption, it is in $\left[\frac{2}{3}, 1\right]$, the domain of $\zeta(\cdot)$. Thus, $\zeta(\cdot)$ is increasing on its domain.

□

Proof of Lemma 1.12 (Marginal Value of Compensation under Equal Altruism and Uniform Valuations).

Assume that $\alpha_i = \alpha_j =: \alpha$, and plug into formula (A.21) for $W'\Big(0; \alpha_i, \alpha_j\Big)$. The result simplifies to

$$W'(0;\alpha,\alpha) = \frac{2(\alpha-1)^2(\alpha+1)}{(\alpha-2)^2(\alpha+2)}.$$

All terms here are nonnegative for $\alpha \in [0, 1]$, and strictly positive for $\alpha < 1$. So, the overall expression is has the same property.

□

Proof of Corollary 1.13 (Relative Sign of Marginal Value of Compensation).

By Proposition 1.11, if $W'\Big(0; \alpha_L, \alpha_j\Big) > 0$, then either $\alpha_j < \frac{2}{3}$, or $\alpha_j \geq \frac{2}{3}$ with $\alpha_L > \zeta\Big(\alpha_j\Big)$

which implies $\alpha_H > \zeta\Big(\alpha_j\Big)$ as well. In either case, $W'\Big(0; \alpha_H, \alpha_j\Big) > 0$.

Similarly, if $W'\Big(0; \alpha_H, \alpha_j\Big) < 0$, then $\alpha_j > \frac{2}{3}$ and $\alpha_H < \zeta\Big(\alpha_j\Big)$, so $\alpha_L < \zeta\Big(\alpha_j\Big)$ also. So, $W'\Big(0; \alpha_L, \alpha_j\Big) < 0$. □

A.2 Proofs for Chapter 2 on Uncertainty in a Trading Partnership

A.2.1 Proofs for Section 2.3 on Equilibrium Properties

Proof of Proposition 2.1 – Non-Autarkic Equilibrium Cutoffs.

The individuals choose whether to trade to maximize their interim expected utility (2.3)–(2.4).[2] As in Chapter 1, this means trading when the interim utility of trading is non-negative, and not trading when it is negative. For the buyer, this means:

$$\begin{aligned}
0 \leq U_b^\theta(v_b, \boldsymbol{\kappa}_s) &= \mathbb{E}_{\varphi\sim\mu_\theta}\left[\int_0^{\kappa_s^\varphi} u_b^\theta(v_b, v_s)\,\mathrm{d}F_s(v_s)\right] \\
&= \mathbb{E}_{\varphi\sim\mu_\theta}\left[\int_0^{\kappa_s^\varphi} (v_b - \alpha_\theta v_s - (1-\alpha_\theta)p)\,\mathrm{d}F_s(v_s)\right]\mathrm{d}F_s(v_s) \\
&= (v_b - (1-\alpha_\theta)p)\cdot\mathbb{E}_{\varphi\sim\mu_\theta}\left[\int_0^{\kappa_s^\varphi}\mathrm{d}F_s(v_s)\right] - \mathbb{E}_{\varphi\sim\mu_\theta}\left[\alpha_\theta\int_0^{\kappa_s^\varphi} v_s\,\mathrm{d}F_s(v_s)\right] \\
&= (v_b - (1-\alpha_\theta)p)\mathbb{E}_{\varphi\sim\mu_\theta}\left[F_s\Big(\kappa_s^\varphi\Big)\right] - \mathbb{E}_{\varphi\sim\mu_\theta}\left[\alpha_\theta\int_0^{\kappa_s^\varphi} v_s\,\mathrm{d}F_s(v_s)\right] \\
\Leftrightarrow v_b &\geq (1-\alpha_\theta)\,p + \alpha_\theta\,\frac{\mathbb{E}_{\varphi\sim\mu_\theta}\left[\alpha_\theta\int_0^{\kappa_s^\varphi} v_s\,\mathrm{d}F_s(v_s)\right]}{\mathbb{E}_{\varphi\sim\mu_\theta}\left[F_s\Big(\kappa_s^\varphi\Big)\right]}.
\end{aligned}$$

2. We assume that they break a tie by trading when utility is zero, but this assumption does not change any results, since this is a zero-probability event.

That is, the best response by a buyer of type θ is to play cutoff

$$\kappa_b^\theta = (1-\alpha_\theta)\,p + \alpha_\theta \cdot \frac{\mathbb{E}_{\varphi\sim\mu_\theta}\left[\int_0^{\kappa_s^\varphi} v_s \,\mathrm{d}F_s(v_s)\right]}{\mathbb{E}_{\varphi\sim\mu_\theta}\left[F_s\left(\kappa_s^\varphi\right)\right]},$$

as was to be shown.

The proof for the seller proceeds analogously.

□

Proof of Proposition 2.3 – Equilibrium Existence with Uncertainty.
If all the buyer types refuse to trade, then all seller types are indifferent among their actions. If they choose to also never trade, then the buyers' action of not trading is in fact a best-response. Thus, autarky is an equilibrium.

For non-autarkic equilibrium existence, note that thanks to (2.9)–(2.10), it is clear that just as without uncertainty, here a player i's best-response cutoff is a weighted average of the price and her partner's expected valuation conditional on trading, with weights $1-\alpha_i$ and α_i, and with partner type distributed based on the updated posterior after observing her partner j trading.

Since the bounding argument used in the Proof of Proposition 1.1 relied only on the support of valuations, it applies here to an average of partner types' valuations just the same. Having an interior price p, or having all types have positive altruism, is enough to ensure that all types will choose to trade with positive probability, regardless of what they think others are playing. So, any solution to the best-response system of equations that has cutoffs in support must have them in the interior of the support, which implies that the equilibrium is non-autarkic. □

A.2.2 Proofs for Section 2.4 on 1×1 *Price Choice*

Proof of Proposition 2.4 – 1×1 *Equilibrium p Choice.*

The Proposition is a direct corollary of Proposition 1.11.

Define the random-role altruism threshold function $\zeta(\cdot)$ as

$$\zeta(\alpha_2) := \frac{-3\alpha_2^2 + 5\alpha_2 - 6 + \sqrt{3}\sqrt{3\alpha_2^4 - 2\alpha_2^3 - 9\alpha_2^2 - 4\alpha_2 + 12}}{2\alpha_2^2 - 6\alpha_2}. \tag{A.22}$$

(This is the same function as (A.19).)

Proposition 1.11 established that when $\alpha_2 \geq \frac{2}{3}$ and $\alpha_1 < \zeta(\alpha_2)$, then $W'(0; \alpha_1, \alpha_2) < 0$. Moreover, because $W(p; \alpha_1, \alpha_2)$ is quadratic in p, having a local max at 0 also means having a global max with in $[0, 1]$.

To be a PBE (Perfect Bayesian Equilibrium), 1 has to choose a p that maximizes the value of the partnership, and 2 has to make an inference about 1's type consistent with that choice. The inference part is vacuously satisfied, since there is only one type of 1 for 2 to assign 100% probability to. The maximization part is satisfied for a choice of p_1 precisely at those parameter values for which $W(p; \alpha_1, \alpha_2)$ is maximized at 0. This proves the Proposition.

□

A.2.3 Proofs for Section 2.5 on 2×1 *Price Choice*

In order to prove Proposition 2.5, we first establish three helpful lemmas about how the value of the partnership depends on 1's reputation.

Lemma A.2 (Value of altruistic reputation at $p = 0$)**.**

Suppose there are two types L and H of player 1, with $\alpha_L < \alpha_H$, and one type of player 2. Then, the value of the partnership to either type at $p = 0$ is higher with the more altruistic pure reputation than with a the more selfish one. That is, $W^\theta\big(0 \mid \alpha_2, \mathbb{1}_H\big) > W^\theta\big(0 \mid \alpha_2, \mathbb{1}_L\big)$ for $\theta \in \{L, H\}$.

Proof of Lemma A.2 – Value of altruistic reputation at $p = 0$.
To show that the value of an altruistic reputation is positive at $p = 0$, we solve for value of the partnership to a player α_θ, facing a player 2 (of altruism α_2) who holds belief $\mathbb{1}_\varphi$ (i.e. is certain that 1's altruism is some α_φ); then we compare this value when $\alpha_\varphi = \alpha_H$ vs. when $\alpha_\varphi = \alpha_L$.

This value is straightforward to compute: first solve for the equilibrium strategies in a 1×1 game with altruism parameters α_2 and α_φ using (1.14)–(1.15), then compute θ's best-response to 2's strategy using (1.12)–(1.13), and then plug into the utility functions (1.10)–(1.11) and (1.19). This value of the partnership—for general p and more specifically at $p = 0$—comes out to

$$\begin{aligned} W^\theta\big(p \mid \alpha_2, \mathbb{1}_\varphi\big) = \frac{1}{2\left(\alpha_2\alpha_\varphi - 4\right)^3} \Bigg(& \Big(\alpha_2\left(p\alpha_\varphi + p - 1\right) - 2p\Big) \\ & \cdot \Big(\alpha_\theta\left((p-1)\alpha_2 + 2p\right) + (p-1)\left(\alpha_2\alpha_\varphi - 4\right)\Big)^2 \\ & - \Big(\alpha_2\left((p-1)\alpha_\varphi + p\right) - 2p + 2\Big) \\ & \cdot \Big(\alpha_\theta\left(p\alpha_2 + 2p - 2\right) + p\left(\alpha_2\alpha_\varphi - 4\right)\Big)^2 \Bigg) \end{aligned}$$

$$W^\theta\big(0 \mid \alpha_2, \mathbb{1}_\varphi\big) = \frac{4(\alpha_\theta)^2\left(\alpha_2\alpha_\varphi - 2\right) - \alpha_2\left(\alpha_\theta\alpha_2 + \alpha_2\alpha_\varphi - 4\right)^2}{2\left(\alpha_2\alpha_\varphi - 4\right)^3}. \tag{A.23}$$

We wish to show that when we successively plug in $\alpha_\varphi = \alpha_H$ and $\alpha_\varphi = \alpha_L$, the difference is positive; that is, that $W^\theta\big(0 \mid \alpha_2, \mathbb{1}_H\big) - W^\theta\big(0 \mid \alpha_2, \mathbb{1}_L\big) > 0$. To that end, differentiate

(A.23):

$$\frac{d}{d\alpha_\varphi}W^\theta\big(0 \mid \alpha_2, \mathbb{1}_\varphi\big) = \frac{1}{2\left(\alpha_2\alpha_\varphi - 4\right)^4} \cdot \alpha_2 \cdot \Bigg((\alpha_\theta)^2 \left(3\alpha_2^3 - 8\alpha_2\alpha_\varphi + 8\right) + 4\alpha_\theta\alpha_2^2\left(\alpha_2\alpha_\varphi - 4\right) + \alpha_2\left(\alpha_2\alpha_\varphi - 4\right)^2 \Bigg).$$

This is in fact positive, and we show it by signing its components. First, the denominator is positive. Second, the α_2 in front upstairs is positive too. What's left to sign is

$$(\alpha_\theta)^2 \left(3\alpha_2^3 - 8\alpha_2\alpha_\varphi + 8\right) + 4\alpha_\theta\alpha_2^2\left(\alpha_2\alpha_\varphi - 4\right) + \alpha_2\left(\alpha_2\alpha_\varphi - 4\right)^2,$$

and it can be arranged as

$$(\alpha_2\alpha_\varphi - 4) \cdot \left[-2\alpha_\theta^2 + \alpha_2(\alpha_2\alpha_\varphi - 4) + 4\alpha_\theta\alpha_2^2\right] + 3\alpha_\theta^2\alpha_2^2.$$

This is positive if the bracketed piece is negative. That piece can be written as

$$-2\alpha_\theta^2 + 4\alpha_2(-1 + \alpha_2 \cdot (\alpha_\theta + \frac{1}{4}\alpha_\varphi)),$$

and if it were positive at any parameter values, then it would also be positive at $\alpha_2 = \alpha_\varphi = 1$ (since the part in round parentheses is increasing in both α_2 and α_φ, and so increasing either of those to 1 would only keep it positive). Subbing in 1 for these parameters turns the expression into

$$-2\alpha_\theta^2 - 4 + 4\alpha_\theta + 1,$$

which more simply is

$$-2\alpha_\theta^2 + 4\alpha_\theta - 3,$$

a quadratic expression that attains its max of -1 at $\alpha_\theta = 1$, and so negative for all $\alpha_\theta \in \mathbb{R}$.

This establishes that $\frac{d}{d\alpha_\varphi}W^\theta\big(0 \mid \alpha_2, \mathbb{1}_\varphi\big) > 0$ for all α_θ and all α_φ. Hence the difference in this function over a non-infinitesimal increase in reputation, $W^\theta\big(0 \mid \alpha_2, \mathbb{1}_H\big) - W^\theta\big(0 \mid \alpha_2, \mathbb{1}_L\big)$

(which goes from $\mathbb{1}_\varphi = \mathbb{1}_L$ to $\mathbb{1}_\varphi = \mathbb{1}_H$), is > 0 as well, as was to be shown.

□

Lemma A.3 (Extreme-belief lower bounds to value of the partnership at $p = \frac{1}{2}$)**.** *Suppose there are two types of player* 1*, with* $\alpha_H > \alpha_L$*, and one type of player* 2*. Then the value of the partnership to* H *at* $p = \frac{1}{2}$ *across all possible beliefs* μ_2 *is minimized at a pure belief. That is,* $W^H\left(\frac{1}{2} \mid \mu_2\right) \geq \min\left\{W^H\left(\frac{1}{2} \mid \mathbb{1}_L\right), W^H\left(\frac{1}{2} \mid \mathbb{1}_H\right)\right\}$ *for all beliefs* μ_2 *about* 1*'s type.*

Proof of Lemma A.3 – Extreme-belief lower bounds to value of the partnership at $p = \frac{1}{2}$*.*

We will prove the Lemma by analyzing $W^H\left(\frac{1}{2} \mid \mu_2\right)$ as a function of 2's belief μ_2 about 1's type (represented as $\mu_2(H) \in [0, 1]$, the probability that she assigns to 1 being the high type). We will show that $W^H\left(\frac{1}{2} \mid \mu_2\right)$, has no interior min for $\mu_2(H) \in [0, 1]$, and so its minimum must occur at one of its endpoints, either 0 or 1. To do this, we will first show that any critical point of $W^H\left(\frac{1}{2} \mid \mu_2\right)$ in μ_2 must also be a critical point of ex-ante utility $\mathcal{U}_s^H\left(\kappa_s^H, \kappa_b^2; p\right)$ in κ_b^2 (where κ_s^H is set to H's best-response to κ_b^2), and that this function can have at most one such critical point. Second, we will show that $W^H\left(\frac{1}{2} \mid \mu_2\right)$ is decreasing in $\mu_2(H)$ at $\mu_2(H) = 1$, and so any interior critical point must be a max.

At the outset, we restrict attention to one role assignment, where 1 is the seller and 2 is the buyer. This is without loss of generality at $p = \frac{1}{2}$, because the best-response equations are symmetric about $\frac{1}{2}$ in p and κ (the equations still hold if we swap p to $1 - p$, which for $p = \frac{1}{2}$ is no change at all, and swap all strategies κ with $1 - \kappa$).So $W^\theta\left(\frac{1}{2} \mid \mu_2\right)$ is equivalent to both $W_b^\theta\left(\frac{1}{2} \mid \mu_2\right)$ and $W_s^\theta\left(\frac{1}{2} \mid \mu_2\right)$.

Since H has no uncertainty about 2's type, the value of the partnership to her, based on

(2.6) and (1.11), is

$$W_s^H\big(p \mid \mu_2\big) = \mathcal{U}_s^H\Big(\kappa_s^H, \kappa_b^2; p\Big) \\ = \left(-\frac{0+\kappa_s^H}{2} + \alpha_H\,\frac{1+\kappa_b^2}{2} + (1-\alpha_H)\,p\right)\cdot(1-\kappa_b^2)\cdot\kappa_s^H, \qquad \text{(A.24)}$$

where κ_s^H and κ_b^2 are equilibrium cutoffs given belief μ_2. That equilibrium must satisfy first order conditions given by (2.11)–(2.12), which applied to this 2×1 type space, are:

$$\begin{aligned}
\kappa_b^2 &= (1-\alpha_2)\cdot p + \alpha_2 \cdot \frac{\sum_{\theta\in\{L,H\}}\mu_2(\theta)\,\kappa_s^\theta\cdot\frac{0+\kappa_s^\theta}{2}}{\sum_{\theta\in\{L,H\}}\mu_2(\theta)\,\kappa_s^\theta} \\
\kappa_s^H &= (1-\alpha_H)\,p + \alpha_H\,\frac{1+\kappa_b^2}{2} \\
\kappa_s^L &= (1-\alpha_L)\,p + \alpha_L\,\frac{1+\kappa_b^2}{2}.
\end{aligned}$$

Note that both in these first order conditions, and in ex-ante utility formula (A.24), the only place μ_2 appears is in 2's best-response. In other words, there is only one channel through which belief μ_2 impacts H's value of the partnership.

To highlight the way that $W_s^H\big(p \mid \mu_2\big)$ depends on 2's belief μ_2, we explicitly set H's cutoff to be a best-response to κ_b^2, as

$$\kappa_s^H = \beta_s\Big(\kappa_b^2; \alpha_H, p\Big) := \alpha_H\,\frac{1+\kappa_b^2}{2} + (1-\alpha_H)\,p, \qquad \text{(A.25)}$$

and we write

$$W_s^H\big(p \mid \mu_2\big) = \mathcal{U}_s^H\Big(\beta_s\Big(\kappa_b^2; \alpha_H, p\Big), \kappa_b^2;\, p\Big)\Big|_{\kappa_b^2=\kappa_b^{2*}(\mu_2;p)},$$

where $\kappa_b^{2*}(\mu_2; p)$ is 2's equilibrium strategy under μ_2 and p.

Since the only dependence of $W_s^H\big(p \mid \mu_2\big)$ on μ_2 is indirect, via 2's equilibrium cutoff, to take the total derivative, we employ the chain rule and arrive at simply

$$\frac{d}{d\mu_2(H)} W_s^H\big(p \mid \mu_2\big) \quad = \quad \frac{d}{d\kappa_b^2}\bigg|_{\kappa_b^2=\kappa_b^{2*}(\mu_2;p)} \mathcal{U}_s^H\Big(\beta_s\Big(\kappa_b^2;\alpha_H,p\Big),\kappa_b^2;\,p\Big) \quad \cdot \quad \frac{d}{d\mu_2(H)}\kappa_b^{2*}(\mu_2;p)\,,$$

where $\kappa_b^{2*}(\mu_2;p)$ is 2's equilibrium cutoff. We have a critical point—where $W_s^H\big(p \mid \mu_2\big)$ is flat in $\mu_2(H)$—only if that value of μ_2 makes one of the two factors zero, i.e. if either

$$0 = \frac{d}{d\kappa_b^2}\mathcal{U}_s^H\Big(\beta_s\Big(\kappa_b^2;\alpha_H,p\Big),\kappa_b^2;\,p\Big) \qquad \text{or} \qquad 0 = \frac{d}{d\mu_2(H)}\kappa_b^{2*}(\mu_2;p)\,.$$

The latter is never zero for $p = \frac{1}{2}$; Lemma A.4 proves this.

We now show that the former occurs for at most one interior κ_b^2.

To take this total derivative, thanks to the Envelope Theorem we take a partial derivative instead, skipping differentiating $\mathcal{U}_s^H\Big(\kappa_s^H,\kappa_b^2;p\Big)$ with respect to κ_s^H. This is applicable because κ_s^H is H's best-response to κ_b^2, meaning that H chooses it to maximize her ex-ante utility $\mathcal{U}_s^H\Big(\kappa_s^H,\kappa_b^2;p\Big)$. When we partially differentiate the expansion of $\mathcal{U}_s^H\Big(\kappa_s^H,\kappa_b^2;p\Big)$ in

(A.24), and then substitute and rearrange, we have

$$\begin{aligned}\left.\frac{\partial \mathcal{U}_s^H\left(\kappa_s^H,\kappa_b^2;p\right)}{\partial \kappa_b^2}\right|_{\kappa_s^H=\beta_s\left(\kappa_b^2\right)} &= -\kappa_s^H\cdot\Bigg(-\frac{\alpha_H}{2}(1-\kappa_b^2)-\frac{0+\kappa_s^H}{2} \\ &\qquad +\alpha_H\,\frac{1+\kappa_b^2}{2}+(1-\alpha_H)\,p\Bigg)\Bigg|_{\kappa_s^H=\beta_s\left(\kappa_b^2\right)} \\ &= -\kappa_s^H\cdot\left(-\frac{\alpha_H}{2}(1-\kappa_b^2)-\frac{0+\kappa_s^H}{2}+\kappa_s^H\right)\Bigg|_{\kappa_s^H=\beta_s\left(\kappa_b^2\right)} \\ &= -\kappa_s^H\cdot\left(-\frac{\alpha_H}{2}(1-\kappa_b^2)+\frac{\kappa_s^H}{2}\right)\Bigg|_{\kappa_s^H=\beta_s\left(\kappa_b^2\right)}.\end{aligned}$$

This shows us that H's equilibrium ex-ante-utility $\mathcal{U}_s^H\left(\kappa_s^H,\kappa_b^2;p\right)\Big|_{\kappa_s^H=\beta_s\left(\kappa_b^2\right)}$ has a critical point in only two situations: if either

$$0=\kappa_s^H \qquad \text{or} \qquad 0=-\frac{\alpha_H}{2}(1-\kappa_b^2)+\frac{\kappa_s^H}{2}.$$

The former condition is autarky, and based on best-response equation, it happens in equilibrium only if $\alpha_H=1$ and $p=0$, so we rule it out since our result pertains only to $p=\frac{1}{2}$. The latter condition may occur for some parameter values; plugging in best-response (A.25) for κ_s^H, the condition comes out to

$$\kappa_b^2=\frac{1}{3}+\frac{2p(-1+\alpha_H)}{3\alpha_H},$$

which for some parameter values is in $(0,1)$, but for others is not. Either way, H's equilibrium ex-ante-utility $\mathcal{U}_s^H\left(\kappa_s^H,\kappa_b^2;p\right)\Big|_{\kappa_s^H=\beta_s\left(\kappa_b^2\right)}$ has at most one interior critical point.

Having established that $W_s^H\big(p\mid\mu_2\big)$ has at most one critical point as $\mu_2(H)$ ranges from 0

to 1, we now show that with $p = \frac{1}{2}$, any critical point must be a max.

Explicitly computing the equilibrium, with an arbitrary belief μ_2, the solution is solvable but long. Setting $p = \frac{1}{2}$ (and noting that $W_s^H\left(\frac{1}{2} \mid \mu_2\right) = W^H\left(\frac{1}{2} \mid \mu_2\right)$ as before), and then differentiating with respect to $\mu_2(H)$, and lastly setting $\mu_2(H) = 1$, we arrive at a fairly simple expression:

$$\frac{d}{d\mu_2(H)} W^H\left(\frac{1}{2} \mid \mu_2\right)\Big|_{\mu_2(H)=1} = \\ -\frac{(\alpha_H - 2)(\alpha_2 - 2)\alpha_2(\alpha_H\alpha_2 - 1)(\alpha_H - \alpha_L)\left(\alpha_H\alpha_2 + (\alpha_2 - 2)\alpha_L - 4\right)}{4(\alpha_H\alpha_2 - 4)^4}.$$

Inspecting these terms, all can be signed using the facts that all the α values are in $[0, 1]$, and that $\alpha_H > \alpha_L$. Upstairs are 4 negative terms and 2 positive ones, downstairs is a positive term, and in front is a minus sign, so the expression is unambiguously negative.

This means that $W^H\left(\frac{1}{2} \mid \mu_2\right)$ is eventually decreasing in $\mu_2(H)$, and since it has at most one critical point, it is either decreasing over the entire domain $[0, 1]$, or else initially increasing (at $\mu_2(H) = 0$ and then after some peak it is decreasing. In either case, it has no interior min, so it must achieve its minimum at an endpoint $\mu_2(H) = 0$ or $\mu_2(H) = 1$ – just as was to be shown.

□

Next we prove a Lemma A.4, used as part of the above Lemma A.3.

Lemma A.4 (Monotonicity of strategies with respect to reputation)**.**
Suppose there are two types of player 1*, with* $\alpha_H > \alpha_L$*, and one type of player* 2*; further assume that* $p = \frac{1}{2}$*. Then,* 2*'s equilibrium strategy* κ_r^2 *(for either role* $r = b$ *or* $r = s$*) is monotonic in* 2*'s belief* μ_2*, as it ranges from* $\mu_2(H) = 0$ *to* $\mu_2(H) = 1$*. Strategies* $\kappa_{r'}^H$ *and*

$\kappa_{r'}^L$ *are similarly monotonic (for* $r' = s$ *or* $r' = b$*).*

Proof of Lemma A.4 – Monotonicity of strategies with respect to reputation.

We prove the Lemma for the case where 2 is the buyer and 1 is the seller. The method will be to totally differentiate 2's cutoff strategy κ_b with respect to her belief $\mu_2(H)$ that 1 is the high type, and then to show the terms have constant sign.

The first order conditions for an equilibrium, given by (2.11)–(2.12) and applied to this 2×1 type space, are (just as in the proof of Lemma A.4):

$$\begin{aligned}
\kappa_b^2 &= (1-\alpha_2)\cdot p + \alpha_2 \cdot \frac{\sum_{\theta\in\{L,H\}} \mu_2(\theta)\,\kappa_s^\theta \cdot \frac{0+\kappa_s^\theta}{2}}{\sum_{\theta\in\{L,H\}} \mu_2(\theta)\,\kappa_s^\theta} \\
\kappa_s^H &= (1-\alpha_H)\,p + \alpha_H\,\frac{1+\kappa_b^2}{2} \\
\kappa_s^L &= (1-\alpha_L)\,p + \alpha_L\,\frac{1+\kappa_b^2}{2}.
\end{aligned}$$

All three equilibrium strategies depend on belief μ_2, though κ_s^H and κ_s^L only depend on it indirectly, via their best-response to κ_b^2. The total derivatives of these equilibrium strategies with respect to $\mu_2(H)$, using the chain rule (and that $\mu_2(L) = 1 - \mu_2(H)$), satisfy

$$\begin{aligned}
&\frac{d\kappa_b^2}{d\mu_2(H)} = \frac{\partial \kappa_b^2}{\partial \mu_2(H)} + \sum_{\theta\in\{L,H\}} \frac{\partial}{\partial \kappa_s^\theta}\kappa_b^2 \cdot \frac{d}{d\mu_2(H)}\kappa_s^\theta \\
\text{For } \theta \in \{L,H\}, \quad &\frac{d\kappa_s^\theta}{d\mu_2(H)} = \frac{\partial \kappa_s^\theta}{\partial \kappa_b^2}\cdot\frac{d\kappa_b^2}{d\mu_2(H)},
\end{aligned}$$

and combining these implies that

$$\frac{d\kappa_b^2}{d\mu_2(H)} \cdot \left(1 - \sum_{\theta \in \{L,H\}} \frac{\partial \kappa_b^2}{\partial \kappa_s^\theta} \cdot \frac{\partial \kappa_s^\theta}{\partial \kappa_b^2} \right) = \frac{\partial \kappa_b^2}{\partial \mu_2(H)}. \tag{A.26}$$

To show that $\frac{d}{d\mu_2(H)}\kappa_b^2$ is of constant sign, we show that both the LHS multiplier term (in parentheses) and the RHS partial derivative are each of constant sign. On the RHS of (A.26), the derivative is computed to be

$$\frac{\partial \kappa_s^\theta}{\partial \mu_2(H)} = \frac{1}{2}\alpha_2 \cdot (\kappa_s^H - \kappa_s^L) \cdot \frac{\kappa_s^H \kappa_s^L}{\left[\sum_\theta \mu_2(\theta)\, \kappa_s^\theta\right]^2}.$$

This expression has the same sign as $(\kappa_s^H - \kappa_s^L)$, for any μ_2. Based on H's and L's best-response equations above, that difference is

$$\kappa_s^H - \kappa_s^L = \left(\frac{1 + \kappa_b^2}{2} - p \right) \cdot (\alpha_H - \alpha_L),$$

and at $p = \frac{1}{2}$, both expressions are unambiguously positive. (For general p, they may or may not be of constant sign, depending on whether $\kappa_b^2 \gtrless 2p - 1$ does not change with μ_2).

On the LHS of (A.26), relevant partial derivatives are

$$\begin{aligned}
&\text{For } \theta \in \{L, H\}, \quad \frac{\partial \kappa_s^\theta}{\partial \kappa_b^2} = \frac{\alpha_\theta}{2}, \\
&\text{For } \theta \in \{L, H\}, \quad \frac{\partial \kappa_b^2}{\partial \kappa_s^\theta} = \alpha_2 \cdot \mu_2(\theta) \cdot \frac{\sum_\varphi \mu_2(\varphi) \cdot \kappa_s^\varphi (\kappa_s^\theta - \kappa_s^\varphi/2)}{\left[\sum_\varphi \mu_2(\varphi)\, \kappa_s^\varphi\right]^2}.
\end{aligned}$$

Combining these, and re-arranging to get a double-sum in both numerator and denominator,

the LHS multiplier term of (A.26) comes out to:

$$\begin{aligned}\left(1-\sum_{\theta\in\{L,H\}}\frac{\partial\kappa_b^2}{\partial\kappa_s^\theta}\cdot\frac{\partial\kappa_s^\theta}{\partial\kappa_b^2}\right) &= 1-\sum_\theta\frac{\alpha_\theta}{2}\cdot\alpha_2\cdot\mu_2(\theta)\cdot\frac{\sum_\varphi\mu_2(\varphi)\cdot\kappa_s^\varphi(\kappa_s^\theta-\kappa_s^\varphi/2)}{\left[\sum_\varphi\mu_2(\varphi)\,\kappa_s^\varphi\right]^2}\\ &= 1-\alpha_2\cdot\frac{\sum_\theta\mu_2(\theta)\,\frac{\alpha_\theta}{2}\cdot\sum_\varphi\mu_2(\varphi)\cdot\kappa_s^\varphi(\kappa_s^\theta-\kappa_s^\varphi/2)}{\left[\sum_\theta\mu_2(\theta)\,\kappa_s^\theta\right]\cdot\left[\sum_\varphi\mu_2(\varphi)\,\kappa_s^\varphi\right]}\\ &= 1-\alpha_2\cdot\frac{\sum_\theta\sum_\varphi\mu_2(\theta)\,\mu_2(\varphi)\cdot\frac{\alpha_\theta}{2}\cdot\kappa_s^\varphi\cdot(\kappa_s^\theta-\kappa_s^\varphi/2)}{\sum_\theta\sum_\varphi\mu_2(\theta)\,\mu_2(\varphi)\cdot\quad\cdot\kappa_s^\varphi\cdot(\kappa_s^\theta\qquad)}.\end{aligned}$$

Examining this expression closely, we see that both the numerator and denominator are double-expectations as indices θ and φ span the type space $\{L,H\}$. Moreover, for each (θ,φ) pair, the numerator term is smaller than the denominator term, since $\frac{\alpha_\theta}{2}\leq\frac{1}{2}$, and $(\kappa_s^\theta-\kappa_s^\varphi/2)\leq\kappa_s^\theta$. In addition, the coefficient α_2 is ≤ 1. Hence, the entire expression after the "$1-$" is <1, which means that overall the expression is >0.

Returning to (A.26), we have now shown that both the LHS and RHS expressions are positive when $\alpha_H>\alpha_L$ and $p=\frac{1}{2}$. Hence, $\frac{d\kappa_b^2}{d\mu_2(H)}$ is positive as well, meaning that the equilibrium κ_b^2 is increasing in $\mu_2(H)$. In addition, the best-responses κ_s^H and κ_s^L are increasing in κ_b^2 (and not directly affected by μ_2), so they are increasing as well. Both are monotonic in $\mu_2(H)$, as was to be shown.

To show this monotonicity result with the roles reversed, simply note the symmetry of the best-response system means that if $p\mapsto 1-p$ and all cutoffs $\kappa\mapsto 1-\kappa$, then the cutoffs still satisfy the equilibrium conditions. Since at $p=\frac{1}{2}$, the p transformation is simply $p:\frac{1}{2}\mapsto\frac{1}{2}$, so we directly have that in equilibrium $\kappa_s^2=1-\kappa_b^2$, meaning that all the cutoffs with these swapped roles are monotonic (decreasing) in $\mu_2(H)$.

□

Proof of Proposition 2.5 – 2×1 – No separating equilibrium where only H chooses $p=0$.

The Proposition states that there cannot exist a Perfect Bayesian Equilibrium in which the two 1 types separate, with price 0 chosen by just H and some price $\hat{p} > 0$ chosen by just L. To prove the Proposition, we show that the Incentive Compatibility conditions of such a PBE are mutually incompatible.

In such a putative PBE, if 2 observes 1 choosing a price of 0 or $\hat{p}$, she can infer exactly which type 1 is. Hence, 2's posteriors following these two prices are

$$\pi\big(\cdot \mid 0\big) = \mathbb{1}_H \qquad \text{(i.e. } \pi\big(H \mid 0\big) = 1\text{)}$$
$$\pi\big(\cdot \mid \hat{p}\big) = \mathbb{1}_L \qquad \text{(i.e. } \pi\big(L \mid \hat{p}\big) = 1\text{).}$$

At other prices (in particular at $p = \frac{1}{2}$), which may not be chosen by anyone with positive probability in equilibrium, we cannot pin down 2's posterior solely from the fact that there is an equilibrium. Where we argue that a player will want to deviate to such a price, we will need to argue that she would do so no matter what reputation it would entail.

We split the analysis into two cases, based the sign of $W^{L'}\big(0 \mid \mathbb{1}_L\big)$, the marginal value of compensation to L, if 2 held the (accurate) belief that she was the low type.

- In Case 1, when $W^{L'}\big(0 \mid \mathbb{1}_L\big) < 0$, we show that L wishes to deviate from $p = \hat{p}$ to $p = 0$.
- In Case 2, when $W^{L'}\big(0 \mid \mathbb{1}_L\big) > 0$, and we show that H wishes to deviate from $p = 0$ to $p = \frac{1}{2}$, or else L wishes to deviate to $p = 0$

Thus, in neither case is it incentive compatible for both H to stick with 0, and L to stick with $\hat{p}$.

Case 1. First, suppose that $W^{L'}\big(0 \mid \mathbb{1}_L\big) < 0$. We will show the following, to prove that Incentive Compatibility fails for L because L wishes to deviate from $p = \hat{p}$ to $p = 0$:

$$W^L\big(\hat{p} \mid \mathbb{1}_L\big) \leq W^L\big(0 \mid \mathbb{1}_L\big) < W^L\big(0 \mid \mathbb{1}_H\big)\,. \tag{A.27}$$

Start with the first inequality in (A.27), that $W^L\big(\hat{p} \mid \mathbb{1}_L\big) \leq W^L\big(0 \mid \mathbb{1}_L\big)$. The curve $W^L\big(p \mid \mathbb{1}_L\big)$ represents the value of the partnership to L in a 1×1 game, and so it is quadratic in p and symmetric about $p = \frac{1}{2}$, as shown in Chapter 1, (and again in the proof of Proposition 2.4). So, since $W^L\big(p \mid \mathbb{1}_L\big)$ is decreasing at $p = 0$ by assumption of Case 1, it has a max at 0 in the domain $P \equiv [0, 1]$. Thus from any $\hat{p} \in [0, 1]$, $W^L\big(\hat{p} \mid \mathbb{1}_L\big) \leq W^L\big(0 \mid \mathbb{1}_L\big)$.

Next, for the second inequality in (A.27), that $W^L\big(0 \mid \mathbb{1}_L\big) < W^L\big(0 \mid \mathbb{1}_H\big)$, we deploy Lemma A.2, which directly proves that making 1's reputation more altruistic—moving from $\mathbb{1}_L$ to $\mathbb{1}_H$—makes the value of the partnership to L higher.

Thus, inequality (A.27) holds, as was to be shown. This tells us that if $W^{L'}\big(0 \mid \mathbb{1}_H\big) < 0$, then L prefers to mimic H by choosing $p = 0$. In this case, we cannot have the separating PBE with H separating by choosing a price of 0, because L would deviate from her supposed equilibrium strategy.

Case 2. We now turn to the second case, with $W^{L'}\big(0 \mid \mathbb{1}_L\big) > 0$, and we show that either L wishes to deviate from $\hat{p}$ to 0, or else H would prefer to deviate from $p = 0$ to $p = \frac{1}{2}$.

Comparing value of the partnership at price 0 with reputation $\mathbb{1}_H$ against the value at price $\frac{1}{2}$ with reputation $\mathbb{1}_L$, at least one of the following two inequalities must hold: either

$$W^L\big(0 \mid \mathbb{1}_H\big) > W^L\left(\frac{1}{2} \mid \mathbb{1}_L\right)$$

or

$$W^H\big(0 \mid \mathbb{1}_H\big) \leq W^H\left(\frac{1}{2} \mid \mathbb{1}_L\right).$$

This can be seen by solving the equilibrium and comparing the expressions. We find that the closed-form expression from the gain from switching from $\frac{1}{2}$ to 0 under these reputations is

$$W^\theta\big(0 \mid \mathbb{1}_H\big) - W^\theta\left(\frac{1}{2} \mid \mathbb{1}_L\right) = -\frac{2\alpha_\theta^2(2-\alpha_H\alpha_2)}{(\alpha_H\alpha_2-4)^3} - \frac{\alpha_2(-\alpha_H\alpha_2-\alpha_\theta\alpha_2+4)^2}{2(\alpha_H\alpha_2-4)^3} + \frac{\left(\alpha_2\left(\frac{1}{2}-\frac{\alpha_L}{2}\right)+1\right)\left(\alpha_\theta\left(\frac{\alpha_2}{2}-1\right)+\frac{1}{2}(\alpha_2\alpha_L-4)\right)^2}{(\alpha_2\alpha_L-4)^3},$$

and whenever this is nonnegative at $\alpha_\theta = \alpha_H$, it is positive at $\alpha_\theta = \alpha_L$. So, either this difference is positive for L, or else it is negative for H, or both. Given this, we split into two cases one final time.

Case 2a: $W^L\big(0 \mid \mathbb{1}_H\big) > W^L\left(\frac{1}{2} \mid \mathbb{1}_L\right)$**.** The working assumption in Case 2 is that $W^{L'}\big(0 \mid \mathbb{1}_L\big) > 0$. This means $W^L\big(p \mid \mathbb{1}_L\big)$ has its max at $p = \frac{1}{2}$. Therefore, no matter which $\hat{p}$ is chosen by L in the putative equilibrium (which garners reputation $\mathbb{1}_L$), $W^L\big(\hat{p} \mid \mathbb{1}_L\big) \leq W^L\left(\frac{1}{2} \mid \mathbb{1}_L\right)$. Combining this with the inequality of Case 2a, we see that $W^L\big(0 \mid \mathbb{1}_H\big) > W^L\big(\hat{p} \mid \mathbb{1}_L\big)$. This means that L would prefer to deviate from a price $\hat{p}$ (with low reputation) to a price of 0 (mimicking H to get a high reputation), which violates Incentive Compatibility.

Case 2b: $W^H\big(0 \mid \mathbb{1}_H\big) \leq W^H\left(\frac{1}{2} \mid \mathbb{1}_L\right)$**.** In this case, we show that H prefers to deviate from a price of 0 to a price of $\frac{1}{2}$. In a putative separative equilibrium, 2 infers from a price of 0 that 1's type is H; that is $\pi(0) = \frac{1}{2}$. However, a price of $\frac{1}{2}$ may be played by nobody in equilibrium, which means that 2's inference following a price of $\frac{1}{2}$ is unrestricted. We therefore show that H prefers to deviate under $\pi\left(\frac{1}{2}\right) = \mu_2$, for *any* belief μ_2.

Thanks to Lemma A.3, we can show that H would prefer to deviate at any μ_2 so long as she would prefer to deviate at $\mu_2 = \mathbb{1}_L$ and $\mathbb{1}_H$. Lemma A.3 tells us that $W^H\left(\frac{1}{2} \mid \mu_2\right) \geq W^H\left(\frac{1}{2} \mid \mathbb{1}_L\right)$ and $W^H\left(\frac{1}{2} \mid \mu_2\right) \geq W^H\left(\frac{1}{2} \mid \mathbb{1}_H\right)$ for all μ_2, so if $W^H\left(\frac{1}{2} \mid \mu_2\right) > W^H\left(0 \mid \mathbb{1}_H\right)$ holds for both $\mu_2 = \mathbb{1}_L$ and $\mu_2 = \mathbb{1}_H$, then it holds for all μ_2.

We already have that $W^H\left(0 \mid \mathbb{1}_H\right) \leq W^H\left(\frac{1}{2} \mid \mathbb{1}_L\right)$, by assumption of being in Case 2b. Therefore, all that remains is to show that $W^H\left(0 \mid \mathbb{1}_H\right) \leq W^H\left(\frac{1}{2} \mid \mathbb{1}_H\right)$.

This follows by virtue of being in Case 2. The working assumption in Case 2 is that $W^{L'}\left(0 \mid \mathbb{1}_L\right) > 0$. By Proposition 1.9, this assumption implies that $\alpha_L > \zeta(\alpha_2)$ (or $\alpha_2 < \frac{2}{3}$), where the cutoff $\zeta(\cdot)$ is defined in (A.19). Since $\alpha_H > \alpha_L$, we also have $\alpha_H > \zeta(\alpha_2)$ (or again $\alpha_2 < \frac{2}{3}$), which implies $W^{H'}\left(0 \mid \mathbb{1}_H\right) > 0$. Further, because $W^H\left(p \mid \mathbb{1}_H\right)$ is quadratic in p and symmetric about $\frac{1}{2}$, so (once again), its slope at 0 tells us its global structure, namely that it has a max at $p = \frac{1}{2}$. Thus, $W^H\left(\frac{1}{2} \mid \mathbb{1}_H\right) > W^H\left(0 \mid \mathbb{1}_H\right)$.

Combining these, we indeed see that $W^H\left(0 \mid \mathbb{1}_H\right)$, which is the value of the partnership to H in a putative separating equilibrium, is smaller than both $W^H\left(\frac{1}{2} \mid \mu_2\right)$ for $\mu_2 \in \{\mathbb{1}_L, \mathbb{1}_H\}$, which is the value she would get from deviating $p = \frac{1}{2}$ if this garnered either a low or high pure reputation, and then with Lemma A.3 this implies that these are weakly smaller than $W^H\left(\frac{1}{2} \mid \mu_2\right)$ for any μ_2. Therefore, H would prefer to deviate to $p = \frac{1}{2}$, violating her Incentive Compatibility condition.

These cases are exhaustive, and show that either H violates IC by preferring to switch from $p = 0$ to $p = \frac{1}{2}$, or that L violates IC by preferring to switch from $p = \hat{p}$ to $p = 0$. Therefore, there cannot be such a separating equilibrium, with H choosing a price of 0 and L choosing some other price $\hat{p}$.

□

A.2.4 Proofs for Section 2.6 on 2×2 Price Choice

Proof of Proposition 2.7 – 2×2 – Signaling trust with $p = 0$.

To prove Proposition 2.7—to demonstrate that Example 2.8 constitutes a Perfect Bayesian Equilibrium—we first solve for period -2 equilibrium strategies and utility in the limit as $\varepsilon \to 0$, and second check the period-2 incentive compatibility conditions behind the separating behavior by ℓ and h and the resulting updating rules for L and H.

The results of this analysis appear as Figure 7, which plots the equilibrium strategies by all types, and Figure 6, which plots the calculated value of the partnership for ℓ and h. Both depict these expressions under the only two reputations that 1 my hold in this putative equilibrium – perceiving 2 as ℓ for sure (i.e. $\mu_1 = \mathbb{1}_\ell$, so that $\mu_1(h) = 0$), and perceiving 2 as h for sure (i.e. $\mu_1 = \mathbb{1}_h$, so that $\mu_1(h) = 1$).

Solving the period-2 trading equilibrium. We begin with the first-order conditions. Players choose their cutoffs best-responding in a way that makes them indifferent to trading, given the cutoffs they expect their partner to play and given the posterior belief about their partner's type conditional on the partner trading. These are given by equations (2.11)–(2.12). We solve here given the role assignment where 1 is the seller and 2 is buyer; the analysis proceeds similarly given the other assignment. In this case, preferences and beliefs given by $\alpha_L = 0$, $\alpha_H = \alpha_\ell = \alpha_h = 1$, $\mu_h(H) = 1 - \varepsilon$, and $\mu_\ell(H) = \varepsilon$, from Example 2.8.

In addition, for the putative equilibrium in Example 2.8, there are only two possible beliefs that 1 can hold: either $\mathbb{1}_h$, which assigns probability 1 to h, or $\mathbb{1}_\ell$, which assigns probability 1 to ℓ. These make the equations simpler, as H's equation only references one partner type. We first show solving for the case of $\mu_1 = \mathbb{1}_H$, then where $\mu_1 = \mathbb{1}_L$, both in the role assignment where 1 is the seller and 2 the buyer, since we can solve for the other assignment by reflecting the price and all strategies about $\frac{1}{2}$.

Given these parameters, the first order systems of equations becomes

$$\kappa_s^L = 1 \cdot p + 0 \cdot \frac{0 \cdot \left(1 - (\kappa_b^\ell)^2\right)/2 \; + \; 1 \cdot \left(1 - (\kappa_b^h)^2\right)/2}{0 \cdot \left(1 - \kappa_b^\ell\right) \quad + \; 1 \cdot \left(1 - \kappa_b^h\right)}$$

$$\kappa_s^H = 0 \cdot p + 1 \cdot \frac{0 \cdot \left(1 - (\kappa_b^\ell)^2\right)/2 \; + \; 1 \cdot \left(1 - (\kappa_b^h)^2\right)/2}{0 \cdot \left(1 - \kappa_b^\ell\right) \quad + \; 1 \cdot \left(1 - \kappa_b^h\right)}$$

$$\kappa_b^h = 0 \cdot p + 1 \cdot \frac{(1-\varepsilon) \cdot \left(\kappa_s^L\right)^2/2 \; + \; \varepsilon \cdot \left(\kappa_s^H\right)^2/2}{(1-\varepsilon) \cdot \left(\kappa_s^L\right) \quad + \; \varepsilon \cdot \left(\kappa_s^H\right)}$$

$$\kappa_b^\ell = 0 \cdot p + 1 \cdot \frac{\varepsilon \cdot \left(\kappa_s^L\right)^2/2 \; + \; (1-\varepsilon) \cdot \left(\kappa_s^H\right)^2/2}{\varepsilon \cdot \left(\kappa_s^L\right) \quad + \; (1-\varepsilon) \cdot \left(\kappa_s^H\right)}.$$

The easiest player to solve for is L, since she is totally selfish ($\alpha_L = 0$), which immediately implies that her cutoff is $\kappa_s^L = p$. Also, since H is certain she faces h and is totally altruistic, her cutoff is simply $\kappa_s^H = \frac{1+\kappa_b^h}{2}$. Plugging these into h's equation and rearranging yields a second-order polynomial in one variable:

$$0 = (\kappa_b^h)^2 \cdot \frac{3}{8}(1-\varepsilon) + \kappa_b^h \left(\frac{(1-\varepsilon)}{4} + \varepsilon p\right) - \frac{(1-\varepsilon)}{8} - \frac{\varepsilon p^2}{2}.$$

This equation has two roots, of which the "+" is the equilibrium, while the "−" is spurious (yielding a solution < 0 for $p, \varepsilon \in [0,1]$):

$$\frac{-(1-\varepsilon) - 4\varepsilon p \pm 2\sqrt{(1-\varepsilon)^2 + \varepsilon p^2(4\varepsilon + 3(1-\varepsilon)) + 2\varepsilon(1-\varepsilon)p}}{3(1-\varepsilon)}$$

Plugging in this solution for κ_b^h into the others' best-response equations, we arrive at equi-

librium cutoffs:

$$
\begin{aligned}
\kappa_s^{L*}\big(p \mid \mu_1 = \mathbb{1}_H\big) &= p \\
\kappa_s^{H*}\big(p \mid \mu_1 = \mathbb{1}_H\big) &= \frac{+(1-\varepsilon) - 2\varepsilon p + \sqrt{(1-\varepsilon)^2 + \varepsilon p^2(4\varepsilon + 3(1-\varepsilon)) + 2\varepsilon(1-\varepsilon)p}}{3(1-\varepsilon)} \\
\kappa_b^{h*}\big(p \mid \mu_1 = \mathbb{1}_H\big) &= \frac{-(1-\varepsilon) - 4\varepsilon p + 2\sqrt{(1-\varepsilon)^2 + \varepsilon p^2(4\varepsilon + 3(1-\varepsilon)) + 2\varepsilon(1-\varepsilon)p}}{3(1-\varepsilon)} \\
\kappa_b^{\ell*}\big(p \mid \mu_1 = \mathbb{1}_H\big) &= \frac{1}{6(1-\varepsilon)(-(2+p)\varepsilon^2 + (2-3p)\varepsilon(1-\varepsilon) + 3p(1-\varepsilon)^2)} \\
&\quad \cdot \Bigg(2\varepsilon(1-\varepsilon)((1-\varepsilon) - \varepsilon) \\
&\quad + \varepsilon + p\left(4\varepsilon^2 - 5\varepsilon(1-\varepsilon) + 3(1-\varepsilon)^2\right) \\
&\quad + p^2\left(2\varepsilon^3 + 6\varepsilon^2(1-\varepsilon) - 9\varepsilon(1-\varepsilon)^2 + 9(1-\varepsilon)^3\right) \\
&\quad \varepsilon\left(\varepsilon(p+2) + (1-\varepsilon)(3p-2)\right) \\
&\qquad \cdot \sqrt{(1-\varepsilon)^2 + \varepsilon p^2(4\varepsilon + 3(1-\varepsilon)) + 2\varepsilon(1-\varepsilon)p}\Bigg).
\end{aligned}
$$

We are interested in equilibrium strategies and payoffs for ε close to 0 (though not $= 0$). To that end, we take the limit as $\varepsilon \to 0$ of these, noting that the limit may be discontinuous at $p = 0$ or $p = 1$ for some players and roles:[3]

$$
\begin{aligned}
\lim_{\varepsilon\to 0} \kappa_s^{L*}\big(p \mid \mu_1 = \mathbb{1}_h\big) &= p \\
\lim_{\varepsilon\to 0} \kappa_s^{H*}\big(p \mid \mu_1 = \mathbb{1}_h\big) &= \frac{2}{3} \\
\lim_{\varepsilon\to 0} \kappa_b^{h*}\big(p \mid \mu_1 = \mathbb{1}_h\big) &= \frac{1}{3} \\
\lim_{\varepsilon\to 0} \kappa_b^{\ell*}\big(p \mid \mu_1 = \mathbb{1}_h\big) &= \begin{cases} \frac{p}{2} & \text{if } p > 0 \\ \frac{1}{3} & \text{if } p = 0 \end{cases}
\end{aligned}
$$

These strategies are depicted in Figure 7, at $\varepsilon = .005$. Note that strategies are close to linear

3. Here in this scenario with $\mu_1 = \mathbb{1}_h$ there is a discontinuity only for ℓ and only at $p = 0$; since nobody is best-responding to her, this discontinuity does not propagate into others' equilibrium cutoffs.

in p for interior p; just as in equations (2.13) in footnote 22, which depict equilibrium when $\varepsilon = 0$ exactly. However, here with $\varepsilon > 0$, the strategies change drastically for p very close to zero. The reason is that at $p = 0$, L's probability of trading is zero, so ℓ's posterior belief shifts: she infers that 1's type must be H, and so she behaves the same as does h. And elsewhere, for p far from 0, ℓ's posterior belief remains very close to her prior (that 1 is likely L). But at very small p, where L chooses to trade only with a very small probability, ℓ's prior likelihood ratio of 1's type is $\frac{\pi(H|\ell)}{\pi(L|\ell)} = \frac{\varepsilon}{1-\varepsilon}$, and the likelihood ratio of 2's strategy is $\frac{\mathbb{P}[\text{sell}|L]}{\mathbb{P}[\text{sell}|H]} = \frac{\kappa_s^L}{\kappa_s^H} \approx \frac{p}{2/3}$, so only when p is very small (on the order of $\frac{3}{2}\varepsilon$) does ℓ's posterior belief that she is facing L move away from 1. (This is the reason we study equilibrium here with $\varepsilon > 0$ and take limits, rather than at $\varepsilon = 0$, because with ε, no amount of data could overwhelm ℓ's prior belief that she is facing L.)

The analogous calculations, this time with $\mu_1 = \mathbb{1}_\ell$ instead of $\mathbb{1}_h$, yield the following equilibrium cutoffs:

$$\lim_{\varepsilon\to 0} \kappa_s^{L^*}\big(p \mid \mu_1 = \mathbb{1}_\ell\big) = p$$

$$\lim_{\varepsilon\to 0} \kappa_s^{H^*}\big(p \mid \mu_1 = \mathbb{1}_\ell\big) = \begin{cases} \frac{p+2}{4} & \text{if } p > 0 \\ \frac{2}{3} & \text{if } p = 0 \end{cases}$$

$$\lim_{\varepsilon\to 0} \kappa_b^{h^*}\big(p \mid \mu_1 = \mathbb{1}_\ell\big) = \begin{cases} \frac{p+2}{8} & \text{if } p > 0 \\ \frac{1}{3} & \text{if } p = 0 \end{cases}$$

$$\lim_{\varepsilon\to 0} \kappa_b^{\ell^*}\big(p \mid \mu_1 = \mathbb{1}_\ell\big) = \begin{cases} \frac{p}{2} & \text{if } p > 0 \\ \frac{1}{3} & \text{if } p = 0 \end{cases}$$

Using these limits, for the given role assignment, and a similar set of calculations for the other role assignment (or as a shortcut to that calculation, reflecting p and the cutoff across $\frac{1}{2}$, to take advantage of the symmetry of the problem) we find that value of the partnership

(equilibrium ex ante utility) when 1 believes she faces h ($\mu_1 = \mathbb{1}_h$) is given by

$$\begin{aligned}
\lim_{\varepsilon\to 0} W^L\big(p \mid \mu_1 = \mathbb{1}_h\big) &= \frac{1}{6}(2p^2 - 2p + 1) \\
\lim_{\varepsilon\to 0} W^H\big(p \mid \mu_1 = \mathbb{1}_h\big) &= \frac{4}{27} \quad \approx 0.148 \\
\lim_{\varepsilon\to 0} W^h\big(p \mid \mu_1 = \mathbb{1}_h\big) &= \frac{4}{27} \quad \approx 0.148 \\
\lim_{\varepsilon\to 0} W^\ell\big(p \mid \mu_1 = \mathbb{1}_h\big) &= \frac{1}{16}(-5p^2 + 5p + 1).
\end{aligned}$$

And, the similar set of calculations, for the case of $\mu_1 = \mathbb{1}_\ell$, results in a value of the partnership of

$$\begin{aligned}
\lim_{\varepsilon\to 0} W^L\big(p \mid \mu_1 = \mathbb{1}_\ell\big) &= \frac{1}{6}(p^2 - p + 1) \\
\lim_{\varepsilon\to 0} W^H\big(p \mid \mu_1 = \mathbb{1}_\ell\big) &= \begin{cases} \frac{1}{128}\left(-7p^2 + 7p + 17\right) & \text{if } 0 < p < 1 \\ \frac{499}{3456} \quad \approx 0.144 & \text{if } p = 0 \text{ or } p = 1 \end{cases} \\
\lim_{\varepsilon\to 0} W^h\big(p \mid \mu_1 = \mathbb{1}_\ell\big) &= \begin{cases} \frac{1}{1024}\left(-17p^2 + 17p + 147\right) & \text{if } 0 < p < 1 \\ \frac{4073}{27648} \quad \approx 0.147 & \text{if } p = 0 \text{ or } p = 1 \end{cases} \\
\lim_{\varepsilon\to 0} W^\ell\big(p \mid \mu_1 = \mathbb{1}_\ell\big) &= \frac{1}{16}(-5p^2 + 5p + 1).
\end{aligned}$$

These values are depicted in Figure 6, which shows the value of the partnership for a very small ε.

Verifying the proposed period-1 separating equilibrium The proposed period-1 equilibrium in Example 2.8 calls for the two types of 2 to play $p^h = 0$ and $p^\ell = \frac{1}{2}$. They

know that 1's belief about their type will update based on their choice of price: it will be $\mathbb{1}_H$ if they choose $p = 0$, but will be $\mathbb{1}_L$ otherwise. Given this, we must verify the Incentive Compatibility conditions ensuring that 2 will not benefit from deviating to another p:

$$W^\ell\left(\frac{1}{2} \mid \mathbb{1}_L\right) \geq W^\ell\Big(p' \mid \mathbb{1}_L\Big) \text{ for all } p' > 0 \tag{A.28}$$

$$W^\ell\left(\frac{1}{2} \mid \mathbb{1}_L\right) \geq W^\ell\Big(p';\mid \mathbb{1}_H\Big) \text{ for } p' = 0 \tag{A.29}$$

$$W^h\big(0 \mid \mathbb{1}_H\big) \geq W^h\Big(p' \mid \mathbb{1}_L\Big) \text{ for all } p' > 0 \tag{A.30}$$

Using Figure 6, verifying these a straightforward task. We start with ℓ's possible deviations. The Proposed equilibrium strategy in Example 2.8 dictates that ℓ chooses $p^\ell = \frac{1}{2}$. There are two sorts of deviations to consider, to check if she can increase her equilibrium utility. First, in (A.28), if ℓ deviates from $p^\ell = \frac{1}{2}$ to some $p' \neq 0$, then she will keep her reputation $\mathbb{1}_\ell$. Second, in (A.29), if she deviates from $p^\ell = \frac{1}{2}$ to $p' = 0$, then she will earn reputation $\mathbb{1}_h$. The first sort is not incentive compatible, because $p = \frac{1}{2}$ is the peak of the curve $W^\ell\big(p \mid \mathbb{1}_\ell\big)$. The second sort is also not incentive compatible, because $W^\ell\big(0 \mid \mathbb{1}_h\big) \approx .06 < W^\ell\Big(\frac{1}{2} \mid \mathbb{1}_\ell\Big) \approx .14$.

Next, in (A.30), we examine h's possible deviations from choosing $p^h = 0$. If she deviates to anywhere else, she will get reputation $\mathbb{1}_\ell$, rather than $\mathbb{1}_h$. Since $p = \frac{1}{2}$ is the peak of the curve $W^h\big(p \mid \mathbb{1}_\ell\big)$, if she deviates at all she should choose $p = \frac{1}{2}$. However, doing so would drop her value from $W^h\big(0 \mid \mathbb{1}_h\big) \approx .1469$ to $W^h\Big(\frac{1}{2} \mid \mathbb{1}_h\Big) \approx .1468$.

Lastly, we confirm that 1's updating rule of $\pi\big(h \mid p = 0\big) = 1$ and $\pi\big(h \mid p \neq 0\big) = 0$ conforms to Bayes' rule.

Note that since the 2 types separate, Bayes' rule only requires that $\pi\Big(\cdot \mid p = 0\Big) = \mathbb{1}_h$ and that $\pi\Big(\cdot \mid p = \frac{1}{2}\Big) = \mathbb{1}_\ell$, i.e. that the 1 types infer from observing $p = 0$ that 2 is h and from observing $p = \frac{1}{2}$ that 2 is $\theta_j = \ell$. And indeed the updating rule does this. However, inference following any other p is not restricted by Bayes' rule. In this case, equilibrium dictates they

infer from any other p that 2 is ℓ, and belief is not incompatible with 2's behavior.

In the first plot, $W^\ell\big(p \mid \mathbb{1}_\ell\big)$ coincides almost perfectly with $W^\ell\big(p \mid \mathbb{1}_h\big)$, so that the two can barely be distinguished without zooming very far in.[4] There are two sorts of deviations by ℓ to consider, based on the equilibrium updating rule $\pi\big(\cdot \mid p\big)$ laid out in Example 2.8. First, if ℓ deviates from $p^\ell = \frac{1}{2}$ to some $p' \neq 0$, then she will keep her reputation $\mathbb{1}_\ell$. Second, if she deviates from $p^\ell = \frac{1}{2}$ to $p' = 0$, then she will earn reputation $\mathbb{1}_h$. The first sort is not incentive compatible, because $p = \frac{1}{2}$ is the peak of the curve $W^\ell\big(p \mid \mathbb{1}_\ell\big)$. The second sort is also not incentive compatible, because $W^\ell\big(0 \mid \mathbb{1}_h\big) \approx .06 < W^\ell\Big(\frac{1}{2} \mid \mathbb{1}_\ell\Big) \approx .14$.

The second plot shows value of the friendship for h under the two reputations. For h, her reputation matters a great deal, because she strongly believes (with probability .995) that she is facing H, who is very altruistic and hence very responsive to (what she believes to be) 2's actions. Because the equilibrium dictates that 1's inference is $\pi\big(\cdot \mid p\big) = \mathbb{1}_L$ following any $p \neq 0$, if h deviates from $p^h = 0$, she will be branded as the low type ℓ. The figure shows that it is not incentive compatible for her to deviate like this: deviating to $p = \frac{1}{2}$ would drop her value from $W^h\big(0 \mid \mathbb{1}_h\big) \approx .1469$ to $W^h\Big(\frac{1}{2} \mid \mathbb{1}_h\Big) \approx .1468$. And, deviating to any other p would give her even lower value, since $p = \frac{1}{2}$ is the peak of the curve $W^h\big(p \mid \mathbb{1}_h\big)$.

Lastly, to confirm the updating rule, note that since the 2 types separate, Bayes' rule only requires that $\pi\Big(\cdot \mid p = 0\Big) = \mathbb{1}_h$ and that $\pi\Big(\cdot \mid p = \frac{1}{2}\Big) = \mathbb{1}_\ell$, i.e. that the 1 types infer from observing $p = 0$ that 2 is h and from observing $p = \frac{1}{2}$ that 2 is $\theta_j = \ell$. Their inference following any other p is not restricted by Bayes' rule. In this case, equilibrium dictates they infer from any other p that 2 is ℓ, and these beliefs are not incompatible with 2's behavior. □

Proof of Proposition 2.9 – 2 × 2 – Symmetrically signaling trust with $p = 0$.

4. This near-indifference by ℓ to her reputation occurs because L's action barely depends on her belief about 2's type (or action), since she is is almost completely selfish ($\alpha_L = .01$) and applies a very low weight to 2's actions in her best-response equations. Since ℓ very strongly believes she is facing L (with probability .995), she too does not care very much about her reputation.

This proof proceeds along the same lines as the proof of Proposition 2.9, only with Example 2.10 giving the parameters and Figure 9 plotting the values. Again, we solve the equilibrium using a computer algebra system, and plot the value of the partnership. Here, we refer directly to the Figures to verify the Incentive Compatibility inequalities. First, for the case where 2 is the one choosing p:

- ℓ prefers to pick $p^\ell = \frac{1}{2}$, because this is peak of her curve $W^\ell\big(p \mid \mathbb{1}_\ell\big)$ and because switching to $p = 0$ and garnering high reputation $\mathbb{1}_h$ doesn't increase her value either.

- h chooses $p^h = 0$ because if she deviates she gets low reputation $\mathbb{1}_\ell$, but even the highest point ($p = \frac{1}{2}$) on that curve $W^h\big(p \mid \mathbb{1}_\ell\big)$ is worse for her than sticking at $p_{jh} = 0$ and getting value $W^h\big(0 \mid \mathbb{1}_h\big)$

- Both L and H correctly infer from observing $p = 0$ that 2 is h, and from observing $p = \frac{1}{2}$ that 2 is ℓ. At other p, they are free to assume what they may (in this case, that 2 is ℓ).

Second, for the case where 1 is the one choosing p, the argument runs exactly the same (no surprise, because the parameters are totally symmetric), only swapping 1 with 2, L with ℓ, and H with h. □

REFERENCES

Abdulkadiroğlu, Atila and Kyle Bagwell, "The Optimal Chips Mechanism in a Model of Favors," Working Paper 2012.

_ **and** _ , "Trust, Reciprocity, and Favors in Cooperative Relationships," *American Economic Journal: Microeconomics*, 2013, *5* (2), 213–59.

Abreu, Dilip, David Pearce, and Ennio Stacchetti, "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 1990, *58* (5), 1041–63.

Åkesson, Lisa, "Remittances and Relationships: Exchange in Cape Verdean Transnational Families," *Ethnos*, 2011, *76* (3), 326–347.

Arrow, Kenneth J., "Gifts and Exchanges," *Philosophy & Public Affairs*, 1972, *1* (4), 343–362.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Social Preferences and the Response to Incentives: Evidence from Personnel Data," *Quarterly Journal of Economics*, 2005, *120* (3), 917–962.

Becker, Gary S., *A Treatise on the Family*, Harvard University Press, 1991.

_ , "Should the Purchase and Sale of Organs for Transplant Surgery be Permitted?," *Becker-Posner Blog*, January 1 2006. Available at `http://www.becker-posner-blog.com/2006/01/should-the-purchase-and-sale-of-organs-for-transplant-surgery-be-permitted-becker.html`.

Bénabou, Roland and Jean Tirole, "Intrinsic and Extrinsic Motivation," *The Review of Economic Studies*, 2003, *70* (3), 489–520.

_ **and** _ , "Incentives and Prosocial Behavior," *American Economic Review*, 2006, *96* (5), 1652–1678.

Bhaskar, Dhruva and Evan D. Sadler, "Resource Allocation with Positive Externalities," Working Paper, NYU and Harvard 2017. Available at SSRN 2853085.

Bloch, Maurice, "The Symbolism of Money in Imerina," in "Money and the Morality of Exchange," Cambridge University Press, 1989, chapter 7, pp. 165–190.

_ **and Jonathan P. Parry**, "Introduction: Money and the Morality of Exchange," in "Money and the Morality of Exchange," Cambridge University Press, 1989, chapter 1, pp. 1–32.

Camerer, Colin, "Gifts as Economic Signals and Social Symbols," *American Journal of Sociology*, 1988, pp. S180–S214.

Chassang, Sylvain, "Fear of Miscoordination and the Robustness of Cooperation in Dynamic Global Games With Exit," *Econometrica*, 2010, *78* (3), 973–1006.

Chatterjee, Kalyan and William Samuelson, "Bargaining under Incomplete Information," *Operations Research*, 1983, *31* (5), 835–851.

Dalkiran, Nuh Aygun, Moshe Hoffman, Ramamohan Paturi, Daniel Ricketts, and Andrea Vattani, "Common Knowledge and State-Dependent Equilibria," in "Algorithmic Game Theory" Springer 2012, pp. 84–95.

Dubner, Stephen J. and Steven D. Levitt, "The Gift-Card Economy," *The New York Times Magazine*, January 7 2007. Available at: `http://www.nytimes.com/2007/01/07/magazine/07wwln_freak.t.html`.

Duffy, John and Daniela Puzzello, "Gift Exchange versus Monetary Exchange: Theory and Evidence," *American Economic Review*, 2014, *104* (6), 1735—1776.

Ellingsen, Tore and Magnus Johannesson, "Pride and Prejudice: The Human Side of Incentive Theory," *American Economic Review*, 2008, *98* (3), 990–1008.

_ **and** _ , "Conspicuous Generosity," *Journal of Public Economics*, 2011, *95* (9), 1131–1143.

Falk, Armin and Urs Fischbacher, "A Theory of Reciprocity," *Games and Economic Behavior*, 2006, *54* (2), 293–315.

Gneezy, Uri and Aldo Rustichini, "A Fine is a Price," *The Journal of Legal Studies*, 2000, *29* (1), 1–17.

_ **and** _ , "Pay Enough or Don't Pay at All," *The Quarterly Journal of Economics*, 2000, *115* (3), 791–810.

Gul, Faruk and Wolfgang Pesendorfer, "Interdependent Preference Models as a Theory of Intentions," Working Paper, Princeton University 2010.

Harsanyi, John C., "Games with Incomplete Information Played by "Bayesian" Players, I–III. Part I. The Basic Model," *Management science*, 1967, *14* (3), 159–182.

_ , "Games with Incomplete Information Played by "Bayesian" Players, I–III. Part II. Bayesian Equilibrium Points," *Management Science*, 1968, *14* (5), 320–334.

_ , "Games with Incomplete Information Played by "Bayesian" Players, I–III. Part III. The Basic Probability Distribution of the Game," *Management Science*, 1968, *14* (7), 486–502.

Hauser, Christine and Hugo Hopenhayn, "Trading Favors: Optimal Exchange and Forgiveness," Working Paper 88, Collegio Carlo Alberto 2008.

Kalla, Simo J., "Essays in Favor-Trading," Dissertation, University of Pennsylvania 2010.

Kamenica, Emir, "Behavioral Economics and Psychology of Incentives," *Annual Review of Economics*, 2012, *4* (1), 427–452.

Kiyotaki, Nobuhiro and Randall Wright, "On Money as a Medium of Exchange," *Journal of Political Economy*, 1989, *97* (4), 927–54.

Kranton, Rachel E., "Reciprocal Exchange: A Self-Sustaining System," *The American Economic Review*, 1996, pp. 830–851.

Kucuksenel, Serkan, "Behavioral Mechanism Design," *Journal of Public Economic Theory*, 2012, *14* (5), 767–789.

Lagos, Ricardo and Randall Wright, "A Unified Framework for Monetary Theory and Policy Analysis," *Journal of Political Economy*, 2005, *113* (3), 463–484.

List, John and Jason Shogren, "The Deadweight Loss of Christmas: Comment," *American Economic Review*, December 1998, *88* (5), 1350–55.

Mailath, George J. and Larry Samuelson, "Reputations in Repeated Games," in H. Peyton Young and Shmuel Zamir, eds., *Handbook of Game Theory with Economic Applications*, Vol. 4, Elsevier, 2015, chapter 4, pp. 165–238.

Maskin, Eric and Jean Tirole, "The Principal-Agent Relationship with an Informed Principal: The Case of Private Values," *Econometrica*, 1990, *58* (2), 379–409.

_ **and** _ , "The Principal-Agent Relationship with an Informed Principal, II: Common Values," *Econometrica*, 1992, *60* (1), 1–42.

Mauss, Marcel, *The Gift: the Form and Reason for Exchange in Archaic Community*, Routeledge, 2002 (1950). Translated by W. D. Halls.

Möbius, Markus, "Trading Favors," Manuscript, Harvard University and NBER 2001.

Neilson, William S., "The Economics of Favors," *Journal of Economic Behavior & Organization*, 1999, *39* (4), 387–397.

Prendergast, Canice and Lars Stole, "Monetizing Social Exchange," 2001.

_ **and** _ , "The Non-Monetary Nature of Gifts," *European Economic Review*, 2001, *45* (10), 1793–1810.

Roth, Alvin E., *Who Gets What – and Why: The New Economics of Matchmaking and Market Design*, Houghton Mifflin Harcourt, 2015.

Roy, Nilanjan, "Cooperation without Immediate Reciprocity: An Experiment in Favor Exchange," Working Paper, California Institute of Technology 2012.

Sandel, Michael J., *What Money Can't Buy: The Moral Limits of Markets*, Macmillan, 2012.

Slonim, Robert, Carmen Wang, and Ellen Garbarino, "The Market for Blood," *Journal of Economic Perspectives*, 2014, *28* (2), 177–96.

Smith, Adam, *An Inquiry into the Nature and Causes of the Wealth of Nations*, Liberty-Classics, 1981 (1776).

Sobel, Joel, "Interdependent Preferences and Reciprocity," *Journal of Economic Literature*, 2005, pp. 392–436.

_ , "Signaling Games," *Encyclopedia of Complexity and Systems Science*, 2009, *19*, 8125–8139.

Solnick, Sara and David Hemenway, "The Deadweight Loss of Christmas: Comment," *American Economic Review*, December 1996, *86* (5), 1299–1305.

Stokey, Nancy L. and Robert E. Lucas Jr., *Recursive Methods in Economic Dynamics*, Harvard University Press, 1989.

Titmuss, Richard Morris, *The Gift Relationship: From Human Blood to Social Policy*, London: George Alien & Unwin Ltd., 1970.

Waldfogel, Joel, "The Deadweight Loss of Christmas," *American Economic Review*, 1993, *83* (5), 1328–36.

_ , *Scroogenomics: Why You Shouldn't Buy Presents for the Holidays*, Princeton University Press, 2009.