THEMA
théorie économique, modélisation et applications

THEMA Working Paper n°2023-11
CY Cergy Paris Université, France

Moral motivations in sequential buyer-seller interactions with adverse selection

José Ignacio Rivero Wildemauwe

June 2023

CY CERGY PARIS UNIVERSITÉ
cnrs

Moral motivations in sequential buyer-seller interactions with adverse selection

José Ignacio Rivero-Wildemauwe*

June 14, 2023

Click for most recent version

Abstract: I study a bilateral trade setting with asymmetric information, where one side has all the bargaining power and makes a take-it-or-leave-it price offer. Both agents hold a certain degree of Kantian morality and thus care about what would have happened had their actions been adopted by their counterpart. In order to capture this, I implement a Veil-of-Ignorance approach, whereby players are uncertain about their role and are thus forced to submit strategies for both the case where they are the *Buyer* and the *Seller*. More precisely, in the first stage, both agents propose the price at which they would be willing to buy, while in a second stage they decide whether they would accept to sell at the offered price. *Buyer* and *Seller* roles are randomly assigned in the last stage. I consider adverse selection by assuming that the *Seller* is fully informed about the product's quality, while the *Buyer* can only form an expectation about it. I show that when the degree of morality is low, the expected quality necessary to produce efficient equilibria is lower than that required by purely selfish agents and, moreover, it is decreasing in the intensity of the moral concern. I also find a threshold degree of morality above which only efficient equilibria are possible for any expectation about quality. Moral preferences thus mitigate the adverse selection problem and completely eliminate it when sufficiently strong.

JEL codes: D03; D82; D91; C78

Keywords: bilateral trade; sequential; asymmetric information; *homo moralis*; Veil of Ignorance

*THEMA, CY Cergy Paris Université, Cergy-Pontoise, France. jose-ignacio.rivero-wildemauwe@cyu.fr

1 Introduction

Asymmetric information constitutes one of the main sources of inefficiencies in market-based allocation mechanisms (Myerson and Satterthwaite (1983)). Numerous examples of this pervasive issue have been studied in diverse interactions, such as that between creditors and lenders (Stiglitz and Weiss (1981)), investors and entrepreneurs (Leland and Pyle (1977)), insurers and insured (Rothschild and Stiglitz (1976)) or buyers and sellers (Akerlof (1970)). The effects of information asymmetries in general and of adverse selection in particular have been largely analysed through models where agents are only concerned about their own material outcomes. However, there exists substantial evidence showing deviations from purely selfish motivations and a vast literature modelling this behaviour (see Fehr and Fischbacher (2003) or Camerer (2003)).

In this paper, I focus on a bilateral trade setting with asymmetric information, where one side has all the bargaining power and makes a take-it-or-leave-it price offer. I study the effects of individuals caring not only about their material payoff, but also about what that payoff would be if, *hypothetically*, their own actions were to be adopted by the other agents. This is often described as "Kantian morality" in the Behavioural Economics literature, and exactly how much players are concerned about this counter-factual scenario is referred to as the *degree of morality*. Moral concerns have been shown to be empirically relevant, as experimental evidence consistent with their presence is reported in Levine et al. (2020) and Alger and Rivero-Wildemauwe (2023), while Kantian preferences' large out-of-sample predictive power is documented in Miettinen et al. (2020) and Van Leeuwen and Alger (2023).[1] Moreover, the existing theoretical literature shows that this sort of moral reasoning helps to remedy market failures in a host of different settings (see for example Alger and Weibull (2016)). It is therefore natural to wonder how considering agents endowed with this kind of preferences affects efficiency in a bilateral trade context where information is distributed unequally.

The interaction at the heart of my model has a *Buyer* who proposes a price in the first stage and a *Seller* who decides whether or not she is willing to trade at that price, after having learned it. In order to be able to introduce Kantian moral preferences in this context, I employ a "Veil-of-Ignorance" approach, whereby neither of the two agents knows what their role in the game will be, and thus need to submit strategies for both the case where they are the *Buyer* but also the *Seller*.[2] More precisely, I implement this by having both players set prices in

[1] For evidence of deontological thinking, see for example Capraro and Rand (2018), Tappin and Capraro (2018), Bursztyn, Fiorin, Gottlieb, and Kanz (Bursztyn et al.), Capraro and Vanzo (2019), Bilancini et al. (2020) and Sutter et al. (2020). For the theory grounding these particular moral motivations in evolutionary processes, refer to Alger and Weibull (2013) and Alger et al. (2020).

[2] The "Veil-of-Ignorance" idea was introduced by Vickrey (1945), Harsanyi (1953) and Rawls (1957). Similar approaches to the one used here and in Rivero-Wildemauwe (2023) are employed to study interactions between moral agents (see for example Alger and Weibull (2013), Laslier (2023)).

the first stage, deciding whether they would accept such prices in the second stage, and being randomly assigned the *Buyer* or *Seller* role with equal probability in a third stage. Finally, to allow for adverse selection, I also assume that the quality, either high or low, is randomly drawn with a given probability. The *Seller* knows her particular quality, while the *Buyer* only knows the probabilities. Importantly, costs for either quality are below their respective consumer valuation, which means that the efficient result is for both of them to be traded.

The sequential move game analysed here results in a unique equilibrium when populated by agents who only care about maximising their material payoff. Indeed, in equilibrium either the price equals the low quality cost and only low quality items are traded, or the price equals the high quality cost and both qualities are traded. This crucially depends on the probability of the item being of high quality. When it is weakly above a given threshold, both qualities are exchanged while if it is strictly below, the adverse selection outcome ensues (thus replicating the standard result).

Considering agents with a low but positive degree of morality leads to a lower required probability of high quality for efficient equilibria to obtain. This threshold is decreasing in the degree of morality, meaning that more moral agents are able to reach efficient equilibria for lower values of the probability of high quality. I also find that when the probability of high quality is indeed above the threshold, there exists a continuum of prices that can support the efficient equilibria. These prices can be shown to be strictly larger than the low quality cost but weakly lower than the high quality one. Moreover, the range of this interval is increasing in the degree of morality. In turn, when the probability of high quality is below the necessary threshold, there is a unique equilibrium identical to the adverse selection outcome in the model with agents who are solely profit-maximisers. That is, the price equals the low quality cost and only low quality items are exchanged.

Interestingly, there exists a degree of morality above which only efficient equilibria are possible, no matter the probability of high quality assumed. Thus, morality mitigates the inefficiencies stemming from adverse selection and completely eliminates them when moral concerns are intense enough.

This work brings together two different strands of the theoretical literature. The first one concerns bargaining under asymmetric information (Ausubel et al. (2002)), where a general result is that the Coase Theorem (Coase (1960)) does not hold when there exist information frictions. The second one studies the effects of moral concerns on different settings: Alger and Weibull (2013), Alger and Weibull (2017), Sarkisian (2017), Ayoubi and Thurm (2020), Muñoz Sobrado (2022), Ayoubi and Thurm (2023), Juan-Bartroli and Karagozoglu (2023).

Amongst these, the closest to this paper are Alger and Weibull (2013), Juan-Bartroli and Karagozoglu (2023) and Rivero-Wildemauwe (2023). Alger and Weibull (2013) briefly discuss an Ultimatum Game played behind a Veil of Ignorance and find that in the presence of risk aversion, morality pushes individuals to share the surplus in a more egalitarian fashion. In turn, Juan-Bartroli and Karagozoglu (2023) analyse a Divide-the-Dollar game between two (potentially) moral players, documenting that moral concerns ensure that bargaining is efficient and the egalitarian outcome must be always a Nash equilibrium of the game. Finally, Rivero-Wildemauwe (2023) studies a one-shot bilateral trade game with asymmetric information, finding that morality either completely eliminates or at least reduces the scope for inefficient equilibria to take place.

Unlike the aforementioned references, the present work deals with sequential rather than simultaneous decision-making. In this line, Alger and Weibull (2013) implement the Veil-of-Ignorance by having agents post a proposal and acceptance threshold at the same time, thus rendering the game simultaneous. This is more straightforward to see in Juan-Bartroli and Karagozoglu (2023) and Rivero-Wildemauwe (2023), as the interactions considered there are by definition one-shot. This difference is far from trivial, since simultaneous moves in bilateral trade introduce an additional source of inefficiency: coordination failures resulting from agents bidding too "aggressively". Assuming a sequential move game naturally does away with this issue when the game is solved by backwards induction.

Another relevant difference is that the sequential set-up confers full bargaining power to the party that makes the first move. This source of inefficiency is completely muted in simultaneous move models. The current setting thus allows me to study the effects of moral concerns in an adverse selection context while also permitting for one of the parties to have market power (namely, the *Buyer*). Thirdly, this paper focuses on a trade interaction that allows for asymmetric information between the parties, something that is not tackled in either Alger and Weibull (2013) or Juan-Bartroli and Karagozoglu (2023).

It is worth noting that the experimental literature has also engaged with interactions relevant to the bilateral trade context and decisions made behind a Veil of Ignorance. Van Leeuwen and Alger (2023) include amongst their experimental protocols a "mini" Ultimatum Game played behind a Veil of Ignorance, and generally find that Kantian concerns are relevant to explain decisions. In turn, Alger and Rivero-Wildemauwe (2023) use a Veil-of-Ignorance treatment to assess changes in the propensity of sellers to actually sell over-priced lemons when role uncertainty is made salient and also to estimate the degree of morality. In that study, some decisions are made behind the Veil of Ignorance (that is, under role uncertainty) and others in front it (when there is no doubt that the decider has actually been attributed that role). In line with previous works using role-uncertainty treatments (see e.g. Iriberri and Rey-Biel (2011)), they find that this makes subjects behave more pro-socially, which in their context means not selling lemons. They notice that this behaviour is thoroughly in line with the presence of Kantian moral concerns.

Lastly, this paper is also broadly related to the theoretical literature that studies the effects of pro-social and moral preferences on the equilibrium outcomes in a host of strategic interactions (see e.g. Arrow (1973), Becker (1976), Andreoni (1990), Bernheim (1994), Levine (1998), Fehr and Schmidt (1999), Akerlof and Kranton (2000), Benabou and Tirole (2006), Alger and Renault (2007), Ellingsen and Johannesson (2008), Englmaier and Wambach (2010), Dufwenberg et al. (2011)).

The rest of the article is organised as follows. Section 2 presents the general framework and the benchmark bilateral trade game under complete information. Section 3 analyses the equilibria under complete information for *homo oeconomicus* and moral agents. Section 4 presents the asymmetric information model. Section 5 analyses the latter's equilibria for different degrees of morality. Finally, Section 6 provides an overall discussion of my results.

2 Framework

As stated in Section 1, I focus on a sequential interaction between a *Buyer* and a *Seller* where in a first stage the former proposes a price and in the second one the latter decides whether to accept it or not. If the *Seller* agrees, then they exchange the object at the settled-upon price. In this case, the *Seller* obtains a payoff that consists of the price minus her cost r, while the *Buyer* gets her valuation v for the item minus the price. If the *Seller* rejects, the payoffs for both are nil. I assume throughout that $0 \leq r < v$, so that trade produces a net positive surplus.

In order to consider Kantian moral motivations, I assume that agents play this game behind a Veil of Ignorance. That is, they choose their actions without knowing whether in the end they will be the *Buyer* or the *Seller* and thus are obliged to post strategies that indicate their actions in either role.[3] Formally, I implement this by having players simultaneously set prices $p_1 \in \mathbb{R}$ and $p_2 \in \mathbb{R}$ in the first stage. In the second stage, after having learned the prices, they simultaneously make accept or reject decisions, which I denote by $\mathscr{a}_1$ for Player 1 and $\mathscr{a}_2$ for Player 2. I represent i's decision to accept or reject with $\mathscr{a}_i = 1$ and $\mathscr{a}_i = 0$ respectively.

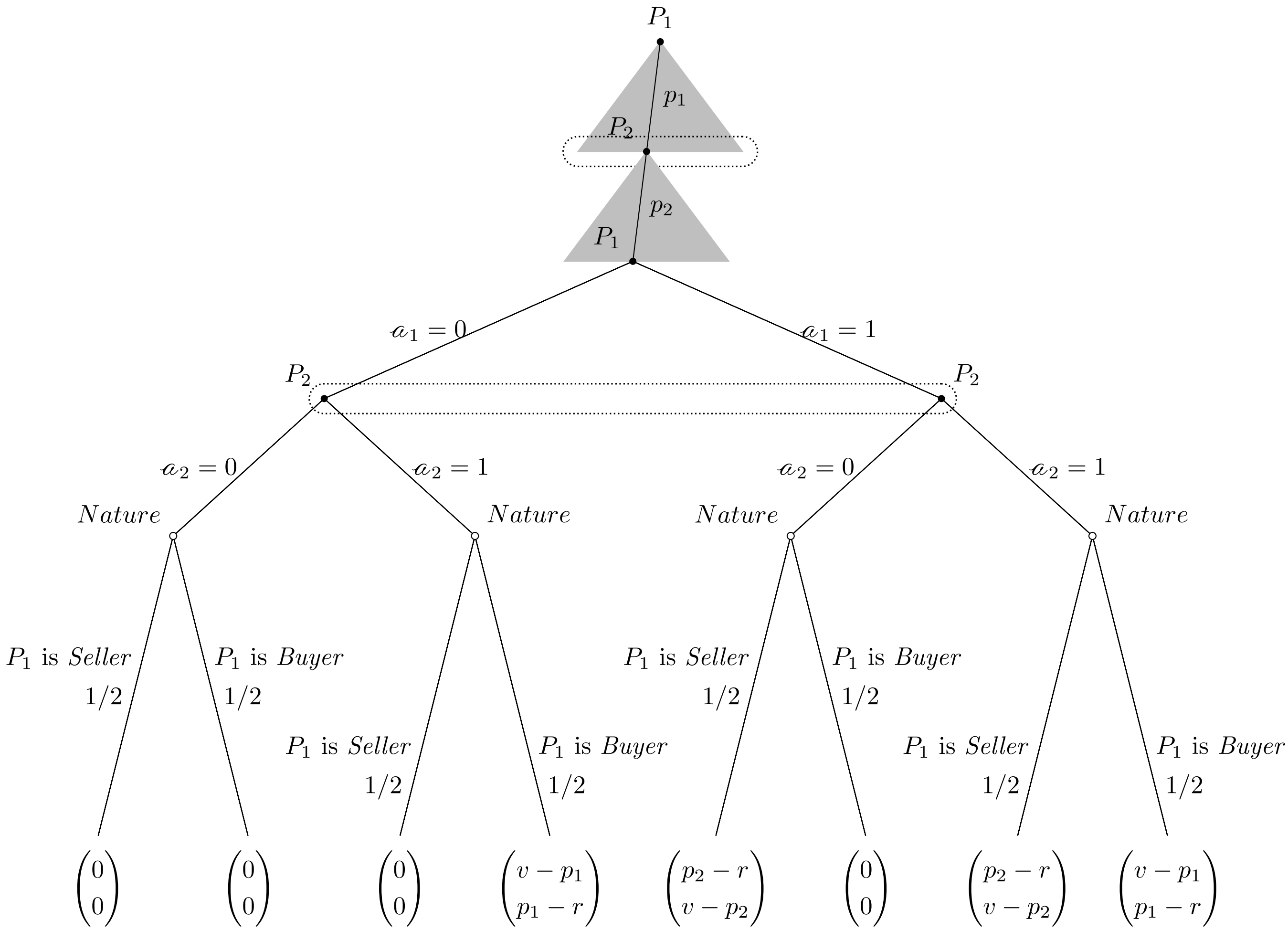
Finally, I introduce a third stage where Nature decides whether Player 1 is the *Buyer* and Player 2 the *Seller* or vice-versa. Either role distribution is chosen with the same probability. Notice that the move by Nature introduced in stage three effectively "blinds" agents with respect to their role and forces them to think about their choices in both roles. I present in Figure 1 the complete information extensive-form game (which I modify in Section 4 to consider asymmetric information).

2.1 Accept or reject decisions

To accommodate potential moral concerns, the accept or reject decision for Player i, $\mathscr{a}_i, i \in \{1, 2\}$, must originate from a mapping of both the price she is pondering about (call it $p \in \mathbb{R}$) and her own price $p_i \in \mathbb{R}$ into $\{0, 1\}$. We thus have $a_i : \mathbb{R}^2 \to \{0, 1\}$. The expression $a_i(p, p_i) = 1$ reflects Player i's decision to accept price p given that she herself is setting p_i. Likewise, $a_i(p, p_i) = 0$ indicates Player i's rejection of price p given that she is setting p_i. I define the space of all possible a_i mappings as $\mathscr{A}$. We will see further on that a_i's second argument p_i matters only in a game played between moral agents.

[3]The present one is adaptation of the approach used in Rivero-Wildemauwe (2023) to be able to accommodate sequential decision-making.

Figure 1: Complete information game in extensive form



2.2 Strategy spaces and payoffs

Given that players do not know their role *ex-ante*, the expected material payoff obtained by Player $i \in \{1,2\}$ when choosing actions $(p_i, \cancel{a}_i) \in \mathbb{R} \times \{0,1\}$ against Player j's $(p_j, \cancel{a}_j) \in \mathbb{R} \times \{0,1\}$ (with $i \neq j$) is:

$$\pi\left((p_i, \cancel{a}_i), (p_j, \cancel{a}_j)\right) = \frac{\cancel{a}_j \cdot (v - p_i) + \cancel{a}_i \cdot (p_j - r)}{2}. \tag{1}$$

Following Alger and Weibull (2013) and the subsequent literature on Kantian moral concerns, I now introduce the utility function for an *homo moralis*:

$$\begin{aligned} U\big((p_i, a_i), (p_j, a_j)\big) =& (1-\kappa) \cdot \pi\big((p_i, a_j(p_i, p_j)), (p_j, a_i(p_j, p_i))\big) + \\ & \kappa \cdot \pi\big((p_i, a_i(p_i, p_i)), (p_i, a_i(p_i, p_i)\big), \end{aligned} \tag{2}$$

where $\kappa \in [0,1]$ is the "degree of morality". This function states that a (partially) moral agent puts a weight of $(1-\kappa)$ on their own expected material payoff when playing (p_i, a_i) against (p_j, a_j). At the same time, she also cares (with a weight of κ) about what her payoff would have been had the other agent used her own strategy (p_i, a_i) instead of (p_j, a_j). This latter term reflects the player's Kantian moral concern, as it effectively answers the question "how would I like it for the other agent to behave like me?". Replacing the payoffs, we get:

$$\begin{aligned} 2U\big((p_i, a_i), (p_j, a_j)\big) =& (1-\kappa) \cdot \Big[a_j(p_i, p_j) \cdot (v - p_i) + a_i(p_j, p_i) \cdot (p_j - r)\Big] + \\ & \kappa \cdot \Big[a_i(p_i, p_i) \cdot (v - p_i) + a_i(p_i, p_i) \cdot (p_i - r)\Big] = \\ & (1-\kappa) \cdot \Big[a_j(p_i, p_j) \cdot (v - p_i) + a_i(p_j, p_i) \cdot (p_j - r)\Big] + \\ & \kappa \cdot \Big[a_i(p_i, p_i) \cdot (v - r)\Big] \end{aligned} \tag{3}$$

From expression (3), it is clear that in addition to their expected material payoff, moral agents value the fact that their price and acceptance decisions lead to surplus-generating trade taking place if their actions are adopted by their counterpart. This is reflected in the second term of the utility function.

2.3 Equilibrium conditions

In the case where agents are endowed with moral preferences, considering a_i mappings that take both prices into account is imperative. To see this, it suffices with examining Equation (3) and noticing that its second term indicates that agent i actually has to decide whether or not she would accept her own price. But the subgame where i sets p_i is not the same where she decides whether to accept p_j. In other words, a player's utility function within a given subgame is affected by actions chosen in another, counter-factual subgame (this is due to the non-consequentialistic nature of Kantian thinking). However, notice that this effect is constant within the subgame and, hence, does not affect the optimal action in it. We can thus solve the game by backwards induction as usual, but using an equilibrium notion that takes non-consequalism into account. Our equilibrium concept must also boil down to standard subgame perfection when dealing with fully consequentialistic agents such as *homo oeconomicus*.

Definition 2.1 (Equilibrium with complete information)**.** *An equilibrium in this game is a profile $(p_i^*, a_i^*) \in \mathbb{R} \times \mathscr{A}$ for all $i \in \{1, 2\}$ such that for any pair of prices $(p_i, p_j) \in \mathbb{R}^2$:*

1. *For all p,*

$$a_i^*(p, p) \in \underset{\mathscr{a} \in \{0,1\}}{\arg\min} (1-\kappa) \cdot \pi\left((p, \mathscr{a}), (p, a_j^*(p, p))\right) + \kappa \cdot \pi\left((p, \mathscr{a}), (p, \mathscr{a})\right) \tag{4}$$

2. *For all (p_i, p_j) such that $p_i \neq p_j$,*

$$a_i^*(p_i, p_j) \in \underset{\mathscr{a} \in \{0,1\}}{\arg\min} (1-\kappa) \cdot \pi\left((p_i, \mathscr{a}), (p_j, a_j^*(p_i, p_j))\right) + \kappa \cdot \pi\left((p_i, a_i^*(p_i, p_i)), (p_i, a_i^*(p_i, p_i))\right) \tag{5}$$

And:

$$p_i^* \in \underset{p \in \mathbb{R}}{\arg\min}\ (1-\kappa) \cdot \Big[a_j^*(p, p_j^*) \cdot (v - p) + a_i^*(p_j^*, p) \cdot (p_j^* - r)\Big] + \kappa \cdot \Big[a_i^*(p, p) \cdot (v - r)\Big] \tag{6}$$

Definition 2.1 states that for all $i \in \{1,2\}$ the equilibrium mappings a_i^* should induce accept or reject decisions $\not{a}_i$ that maximise the agents' utility given the prices set in the first stage. In turn, for all $i \in \{1,2\}$ the equilibrium prices p_i^* should constitute a mutual Best Reply given that agents anticipate that the optimal mappings will be chosen.

This Definition retains the subgame perfection "flavour" while allowing to consider agents concerned with the question: "how would I fare if my counterpart behaved like me?". Notice that this reasoning implies that off-equilibrium actions may affect equilibrium utilities. It also suggests a way to find such equilibria: first assume that both p_1 and p_2 are known and compute $a_i^*, i \in \{1,2\}$, at the last stage. Next, in the previous stage and knowing the a_i^* mappings, find $p_i^{BR}(p_j)$ such that it maximises (3).

3 Equilibrium with complete information

3.1 *Homo oeconomicus* ($\kappa = 0$)

In the *homo oeconomicus* case, all the agent cares about is $\pi\left((p_i, \not{a}_i), (p_j, \not{a}_j)\right)$. I apply backwards induction and therefore start with the last stage, where agents need to decide whether to sell or not at the proposed price. It is clear that $\not{a}_i = 1$ if and only if $p_j \geq r$ and 0 otherwise, for every $p_i \in \mathbb{R}$. Given this, in the price-setting stage agents will chose $p_i = r$. We can thus conclude that the complete information game when $\kappa = 0$ has a unique equilibrium: $p_i^* = r$ and $a_i^*(p_j^*, p_i^*) = 1$ if and only if $p_j \geq r$ and 0 otherwise (which given p_i^* induces $\not{a}_i = 1$); for all $i, j \in \{1,2\}, i \neq j$.

In equilibrium, two agents who have absolutely no moral concerns will offer and accept prices equal to cost. Trade thus takes place, with the player in the *Buyer* role capturing all the surplus generated. This is consistent with a standard Ultimatum Game with no role uncertainty. In this line, notice that the *homo oeconomicus* agent only considers whether the price she is being offered is above her cost or not. This means that we can restrict attention to a_i mappings that just link a price with an accept or reject decision, such that $a_i : \mathbb{R} \to \{0,1\}$. It follows that if we only consider $a_i(p_j)$ and not $a_i(p_j, p_i)$, the equilibrium profile found is also the unique subgame-perfect equilibrium.

3.2 *Homo moralis* ($\kappa \in (0,1)$)

When considering moral agents, we need to consider that agents will take their own price into account when deciding whether to accept or reject her rival's. Following the method suggested by Definition 2.1, I employ backwards induction in order to find the game's equilibria. I start with the accept or rejection stage. If $p_i = p_j = p$, utility becomes:

$$
\begin{aligned}
2U\big((p,a_i),(p,a_j)\big) =&(1-\kappa)\cdot\Big[a_j(p,p)\cdot(v-p)+a_i(p,p)\cdot(p-r)\Big]+\\
&\kappa\cdot\Big[a_i(p,p)\cdot(v-r)\Big]=\\
&(1-\kappa)\cdot(a_j(p,p)\cdot(v-p))+a_i(p,p)\cdot((1-\kappa)p+\kappa v-r).
\end{aligned}
\tag{7}
$$

The optimal accept decision a_i thus becomes: $a_i = 1$ if and only if $p \geq \frac{r-\kappa v}{1-\kappa}$ and 0 otherwise. In contrast, if $p_i \neq p_j$, then $a_i = 1$ for all $p_i \neq p_j$ and $a_i^*(p_j, p_i) = 1$ if and only if $p_j \geq r$ and 0 otherwise.

We now need to find p_i such that it is a best response to p_j given the above a_i^* mappings. If $p_j \geq r$, by setting $p_i = r$, Player i ensures that she is buying at the lowest possible price while also capturing the "moral benefit", since $a_i^* = 1$ both for $p_j = r$ and for $p_j > r$. Next, for $p_j \in [\frac{r-\kappa v}{1-\kappa}, r)$, i is indifferent between setting $p_i = r$ or $p_i = p_j$. Finally, for $p_j < \frac{r-\kappa v}{1-\kappa}$, the best response is $p_i = r$.

We conclude that there exists a continuum of equilibria: for all $i, j \in \{1,2\}, i \neq j$, where $p_i^* = p_j^* \in [\frac{r-\kappa v}{1-\kappa}, r]$, $a_i^*(p_j^*, p_i^*) = 1$ if and only if $p_j \geq \frac{r-\kappa v}{1-\kappa}$ and 0 otherwise for $p_i^* = p_j^* = p$, and $a_i^*(p_j^*, p_i^*) = 1$ if and only if $p_j^* \geq r$ and 0 otherwise, for $p_i^* \neq p_j^*$. Moral agents can sustain equilibria with trade at prices lower than the cost. How low they are willing to go depends on their degree of morality κ, with larger values of the latter making the agents more willing to accept lower prices.

The intuition behind this result is simple. When a morally concerned agent sets the same price as her counterpart, at the time of making an accept or reject decision she takes into account what the "moral" action would be. Given that trade generates a net positive surplus, morality pushes the agent to accept. However, she will not be willing to trade at any price because she also cares about her own material surplus. It is from solving this trade-off that equilibrium prices originate, with the agent willing to part with her item when in the *Seller* role only for prices weakly above the "moral reservation price" $\frac{r-\kappa v}{1-\kappa}$. A more intense concern for morality will cause the agent to accept lower prices because making decisions that facilitate trade if universalised is relatively more important to her. In fact, an agent who only cares about morality ($\kappa \to 1$) accepts any price. In contrast, players who do not care much about the moral aspect will not be willing to trade at too low a price when compared with the cost r. In the limit, an agent who does not care about morality ($\kappa \to 0$) only accepts prices at or above cost. We thus observe no discontinuities in the equilibrium sets as κ goes from zero to strictly positive, as is the case in most applications studied in Rivero-Wildemauwe (2023).

It is also worth underlining that in this model, both agents' have the same degree of morality and this is common knowledge. Thus, when choosing a price to offer if in the *Buyer* role, players are aware that when prices are symmetric, the *Seller* will agree to any price above the moral reservation price. As a consequence, they take full advantage of this fact and offer the lowest acceptable price possible, which is precisely that moral reservation price and is weakly below the *Seller*'s cost. So far, things look quite bright for the party who holds market power (namely, the *Buyer*) and indeed they are. But this comes with a caveat when we stop thinking about roles and instead focus on *agents*. More precisely, recall that any given player does not learn her role until after decisions are made. In equilibrium, agents fully exploit their market power as *Buyer*, but are at the same time willing to accept these low prices when in the *Seller* role. Moral agents thus offer low prices in equilibrium, but only provided that they themselves would be willing to accept those low prices.

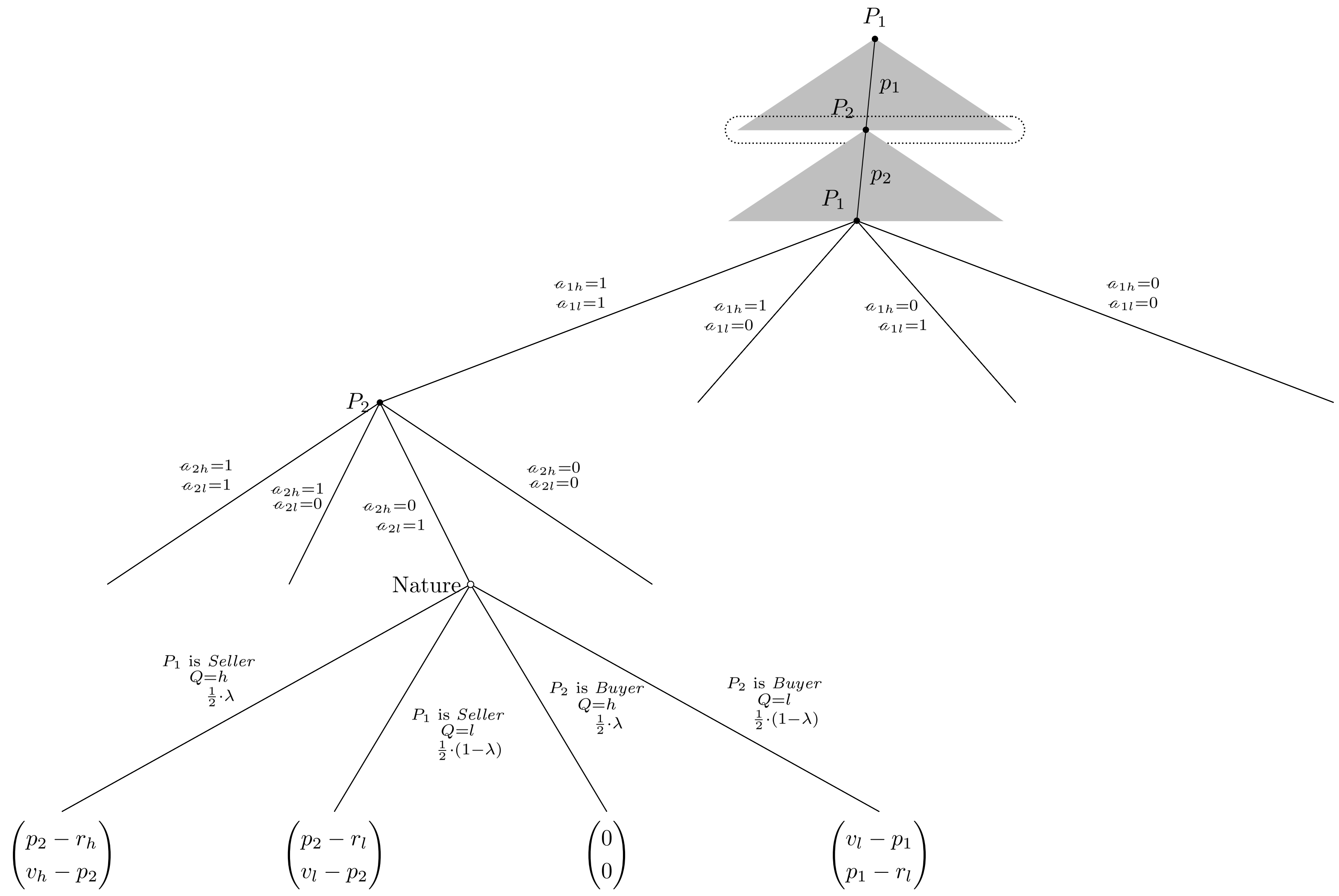
This complete information version of the model resembles an Ultimatum Game played behind a Veil-of-Ignorance. Alger and Weibull (2012) (Section 4) and Alger and Weibull (2013) (Section 6) also study this type of interaction between moral agents, finding that morality essentially pushes individuals to share the surplus in a more egalitarian fashion. However, their modelling approach is different to the one presented here. To begin with, they assume risk averse agents, while in the current paper, payoffs are linear in surplus and thus agents are risk-neutral. In addition, the strategy spaces in their model are different from the ones in the present study, as agents simultaneously post a proposal and acceptance threshold. In contrast, here I am assuming that agents make accept or reject decisions after learning which prices have been posted, thus retaining the sequentiality of the Ultimatum Game at the core of the interaction tackled by Alger and Weibull (2012) and Alger and Weibull (2013).

4 Introducing asymmetric information

I now introduce asymmetric information. The aim is to capture a situation where the *Seller* is aware of her product's quality but the *Buyer* can only form an expectation about it. In order to do that, I assume that even though the *Buyer* can still only set one price, the *Seller* can make two "accept or reject" decisions with regard to that price: one for the case where her product is high quality and one for the case where it is low quality, denoted by $\mathscr{a}_{ih} \in \{0,1\}$ and $\mathscr{a}_{il} \in \{0,1\}$ respectively. In addition, in the game's final stage, Nature draws the the role distribution (with equal probabilities) as well as the item's quality. The latter can be high with probability $\lambda \in [0,1]$ and low with probability $1-\lambda$. In line with this, there is now a valuation for the high quality object and another one for the low quality item, and *idem* for the costs. I denote them by v_h, v_l, r_h and r_l respectively. The modified extensive form game is shown in Figure 2, where the play depicted has Player 1 selling both the high and low quality items, while Player 2 only sells the low quality one.

I assume throughout that $0 \leq r_l < v_l < r_h < v_h$. Notice that the cost of producing each quality r_Q is lower than its respective consumer valuation $v_Q, Q \in \{h,l\}$. This means that the trade of *both* qualities is socially desirable, as there are net positive surpluses to be had from the exchange of either one.

Figure 2: Asymmetric information game in extensive form. In this play, Player 1 is selling both the high and low quality items, while Player 2 only sells the low quality one.



The strategy spaces, payoffs, utility functions and equilibrium conditions must be modified accordingly. The expected material payoff obtained by Player $i \in \{1,2\}$ when choosing actions $(p_i, \cancel{a}_{ih}, \cancel{a}_{il}) \in \mathbb{R} \times \{0,1\}^2$ against Player j's $(p_j, \cancel{a}_{jh}, \cancel{a}_{jl}) \in \mathbb{R} \times \{0,1\}^2$ (with $i \neq j$) is:

$$\begin{aligned}\pi\big((p_i, \cancel{a}_{ih}, \cancel{a}_{il}), (p_j, \cancel{a}_{jh}, \cancel{a}_{jl})\big) =& \frac{1}{2}\Bigg[\lambda \cancel{a}_{jh}(v_h - p_i) + (1-\lambda)\cancel{a}_{jl} \cdot (v_l - p_i)\Bigg] + \\ & \frac{1}{2}\Bigg[\lambda \cancel{a}_{ih} \cdot (p_j - r_h) + (1-\lambda)\cancel{a}_{il} \cdot (p_j - r_l)\Bigg],\end{aligned} \tag{8}$$

The actions $\cancel{a}_{iQ}$ for $Q \in \{h,l\}$ are originated by mappings $a_{iQ} \in \mathscr{A} : \mathbb{R}^2 \to \{0,1\}$ that link the price Player i is considering (p) and her own price p_i to the accept or reject decision. Having re-defined the payoffs, I now show the utility function for an *homo moralis*:

$$\begin{aligned}U\big((p_i, a_{ih}, a_{il}), (p_j, a_{jh}, a_{jl})\big) =& (1-\kappa) \cdot \pi\big((p_i, a_{ih}(p_j,p_i), a_{il}(p_j,p_i), (p_j, a_{jh}(p_i,p_j), a_{jl}(p_i,p_j)\big) + \\ & \kappa \cdot \pi\big((p_i, a_{ih}(p_i,p_i), a_{il}(p_i,p_i), (p_i, a_{ih}(p_i,p_i), a_{il}(p_i,p_i)\big).\end{aligned} \tag{9}$$

Replacing the payoffs, we get:

$$\begin{aligned}2U\big((p_i, a_{ih}, a_{il}), (p_j, a_{jh}, a_{jl}))\big) =& (1-\kappa) \cdot \Big[\lambda a_{jh}(p_i,p_j)(v_h - p_i) + (1-\lambda)a_{jl}(p_i,p_j)(v_l - p_i) + \\ & \lambda a_{ih}(p_j,p_i)(p_j - r_h) + (1-\lambda)a_{il}(p_j,p_i)(p_j - r_l)\Big] + \\ & \kappa \cdot \Big[\lambda a_{ih}(p_i,p_i)(v_h - r_h) + (1-\lambda)a_{ib}(p_i,p_i)(v_l - r_l)\Big].\end{aligned} \tag{10}$$

Definition 4.1 states the equilibrium conditions in the asymmetric information game:

Definition 4.1 (Equilibrium with asymmetric information). *An equilibrium in this game is a profile $(p_i^*, a_{ih}^*, a_{il}^*) \in \mathbb{R} \times \mathscr{A}^2$ for all $i \in \{1,2\}$ such that for any pair of prices $(p_i, p_j) \in \mathbb{R}^2$:*

1. *For all $Q, R \in \{h,l\}, Q \neq R$ and all p,*

$$\begin{aligned}a_{iQ}^*(p,p) \in \underset{\cancel{a}_Q \in \{0,1\}}{\arg\min} \Bigg\{ & (1-\kappa) \cdot \pi\Big((p, \cancel{a}_Q, a_{iR}^*(p,p)), (p, a_{jQ}^*(p,p), a_{jR}^*(p,p))\Big) + \\ & \kappa \cdot \pi\Big((p, \cancel{a}_Q, a_{iR}^*(p,p)), (p, \cancel{a}_Q, a_{iR}^*(p,p))\Big)\Bigg\}\end{aligned} \tag{11}$$

2. *For all $Q, R \in \{h,l\}, Q \neq R$ and all (p_i, p_j) such that $p_i \neq p_j$,*

$$\begin{aligned}a_{iQ}^*(p_i,p_j) \in \underset{\cancel{a}_Q \in \{0,1\}}{\arg\min} \Bigg\{ & (1-\kappa) \cdot \pi\left((p_i, \cancel{a}_Q, a_{iR}^*(p_i,p_j)), (p_j, a_{jQ}^*(p_i,p_j), a_{jR}^*(p_i,p_j))\right) + \\ & \kappa \cdot \pi\left(\left(p_i, a_{iQ}^*(p_i,p_i), a_{iR}^*(p_i,p_i)\right), \left(p_i, a_{iQ}^*(p_i,p_i), a_{iR}^*(p_i,p_i)\right)\right)\Bigg\}\end{aligned} \tag{12}$$

And:

$$p_i^* \in \arg\min_{p\in\mathbb{R}} \Bigg\{ (1-\kappa)\cdot \Big[\lambda a_{jh}^*(p,p_j^*)(v_h - p) + (1-\lambda)a_{jl}^*(p_i,p_j^*)(v_l-p)+$$
$$\lambda a_{ih}^*(p_j,p)(p_j^* - r_h) + (1-\lambda)a_{il}^*(p_j^*,p)(p_j^*-r_l)\Big]+ \tag{13}$$
$$\kappa \cdot \Big[\lambda a_{ih}^*(p,p)(v_h-r_h) + (1-\lambda)a_{ib}^*(p,p)(v_l-r_l)\Big]\Bigg\}.$$

Definition 4.1 states that for all $i \in \{1,2\}$ and all $Q \in \{h,l\}$ the equilibrium mappings a_{iQ}^* should induce accept or reject decisions $\cancel{a}_{iQ}$ that maximise the agents' utility given the prices set in the first stage. In turn, for all $i \in \{1,2\}$ the equilibrium prices p_i^* should constitute a mutual Best Reply given that agents anticipate that the optimal mappings will be chosen.

Turning to the analysis of the game's equilibria under incomplete information, I define the following quantities. Firstly, the expected consumer valuation, which is the average of the high and low quality valuations weighed by their respective probabilities (λ and 1- λ).

$$v_e \equiv \lambda v_h + (1-\lambda)v_l. \tag{14}$$

Secondly, the minimum value of λ for which the expected consumer valuation is weakly above the high-quality cost ($r_h \leq v_e$):

$$\lambda_e \equiv \frac{r_h - v_l}{v_h - v_l}. \tag{15}$$

Next, the function $\lambda_1 : \mathbb{R} \to [0,1]$. It represents the minimum value of λ for which an agent would be willing to pay price p in order to acquire the item without knowing its quality.

$$\lambda_1(p) \equiv \frac{p - r_l}{p - r_l + v_h - r_h}. \tag{16}$$

Recall that in Section 2 a "moral reservation price" sprang out from the computations, and it was equal to $\frac{r-\kappa v}{1-\kappa}$. All prices between this value and the cost r could support an equilibrium with trade. In the asymmetric information case we need to distinguish between the moral reservation price for a high quality from that of a low quality item. Thus:

$$q_Q \equiv \frac{r_Q - \kappa v_Q}{1-\kappa}; Q \in \{h,l\}. \tag{17}$$

It is clear that as the degree of morality κ increases $q_Q, Q \in \{h,l\}$ decrease. It is then useful to compute the values of κ for which moral reservation prices sit below costs and valuations. More exactly, if $\kappa \geq \frac{r_h - r_l}{v_h - r_l}$ then $q_h \leq r_l$ and if $\kappa \geq \frac{r_h - r_l}{v_h - v_l}$ then $q_h \geq q_l$.

$$\kappa_1 \equiv \frac{r_h - r_l}{v_h - r_l}; \kappa_2 \equiv \frac{r_h - r_l}{v_h - v_l}. \tag{18}$$

Notice that given my assumption that $0 \leq r_l < v_l < r_h < v_h$, we have $\kappa_1 < \kappa_2$. Moreover:

1. If $\kappa \leq \kappa_1$, then $\mathbf{q_l} \leq r_l \leq \mathbf{q_h} \leq v_l < r_h < v_h$
2. If $\kappa \in [\kappa_1, \kappa_2]$, then $\mathbf{q_l} \leq \mathbf{q_h} \leq r_l < v_l < r_h < v_h$
3. If $\kappa \geq \kappa_2$, then $\mathbf{q_h} \leq \mathbf{q_l} < r_l < v_l < r_h < v_h$

5 Equilibria with asymmetric information

5.1 *Homo oeconomicus* ($\kappa = 0$)

I now move on to analyse the asymmetric game's equilibria for different degrees of morality. The case where $\kappa = 0$ and therefore agents only care about their own material well-being is tackled in Proposition 5.1.

Proposition 5.1. *Consider the asymmetric information game and assume $\kappa = 0$. Then, the unique equilibrium $(p_i^*, a_{ih}^*, a_{il}^*)$ is $p_i^* = r_h$ and $a_{iQ}^*(p_j^*, p_i^*) = 1 \iff p_j^* \geq r_Q$ and 0 otherwise, for all $i = \{1, 2\}; Q = \{h, l\}$; if and only if $\lambda \geq \lambda_e$. For $\lambda < \lambda_e$, the unique equilibrium $(p_i^*, a_{ih}^*, a_{il}^*)$ is $p_i^* = r_l$ and $a_{iQ}^*(p_j^*, p_i^*) = 1 \iff p_j^* \geq r_Q$ and 0 otherwise, for all $i = \{1, 2\}; Q = \{h, l\}$.*

Proof. Starting from the accept or reject stage, i's payoff (8) is maximised if and only if $a_{iQ} = 1$ when $p_j \geq r_Q$ and 0 otherwise, for $Q \in \{h, l\}$. This is valid *for all* p_i, due to additive separability of the terms depending on p_i and those depending on p_j. The Best Reply mappings are thus $a_{iQ}^{BR}(p_j, p_i) = 1 \iff p_j \geq r_Q$, otherwise $a_{iQ}^{BR}(p_j, p_i) = 0$. At the price-setting stage, the price p_i that maximises i's utility (10) with $\kappa = 0$ is given a_{iQ}^{BR}, a_{jQ}^{BR} and p_j is $p_i^{BR} = r_h$ if $\lambda \geq \lambda_e$ and $p_i^{BR} = r_l$ otherwise. □

Corollary 5.1.1. *Assume that $\kappa = 0$. Since only the first element of the a_i mappings are relevant to compute the Best Replies, payoffs in a given subgame are not affected by decisions made in other subgames. As a consequence, the profiles that are equilibria according to Definition 4.1 are also subgame-perfect equilibria.*

In short, Proposition 5.1 states that with *homo oeconomicus*, we have a unique, symmetric equilibrium where both players trade only low quality at price r_l if $\lambda < \lambda_e$ or both qualities at price r_h when $\lambda \geq \lambda_e$. Recall that the trade of both qualities is socially desirable as it generates a net positive surplus. In this context, when the probability of high quality λ is not large enough, the typical adverse selection result ensues, whereby the high quality items are not traded.

Notice that the equilibria described are not only compliant with Definition 4.1 but are also subgame-perfect, by virtue of the additive separability of the terms depending on p_i and those depending on p_j in expression (8).

5.2 Moral agents ($\kappa \in (0, 1)$)

Moving on to moral agents, the relevant thresholds for the degree of morality are κ_1 and κ_2. When κ surpasses the former, the moral reservation price for the high quality becomes smaller than the low quality cost, leading to both moral reservation prices to sit below both costs. In turn, when κ is larger than κ_2, not only are moral reservation prices lower than costs, but the high quality moral reservation price is below that of the low quality. Proposition 5.2 presents the equilibria for moral agents with a "low" degree of morality, namely, with $\kappa < \kappa_1$.

Proposition 5.2. *Consider the asymmetric information game and assume* $\kappa \in (0, \kappa_1)$*. Then the game's equilibria according to Definition 4.1 are:*

- *When* $\lambda \geq \lambda_1(p)$*, for every* $p \in [q_h, r_h]$ *and every* $\lambda_1(p)$ *:* $p_i^* = p_j^* = p \in [q_h, r_h], a_{iQ}^*(p, p_i^*) = 1 \iff p \geq q_Q$ *and* 0 *otherwise, for all* $Q \in \{h, l\}; i, j \in \{1, 2\}; i \neq j$*.*
- *When* $\lambda < \lambda_1(p)$*, for every* $p \in [q_h, r_h]$ *and every* $\lambda_1(p)$*:* $p_i^* = p_j^* = r_l; a_{iQ}^*(p^*, p_i^*) = 1 \iff p^* \geq q_Q$ *and* 0 *otherwise. for all* $Q \in \{h, l\}; i, j \in \{1, 2\}; i \neq j$*.*

Proposition 5.2 is obtained by applying backwards induction as suggested by Definition 4.1 (for ease of exposition, its proof is relegated to Appendix). It states that if λ is weakly larger than $\lambda_1(p)$ for $p \in [q_h, r_h]$, then there is a continuum of equilibrium prices in the $[q_h, r_h]$ interval. The equilibrium mappings a_{iQ}^* are such that, given their own equilibrium price, agents will accept any price at least as high as q_h. That is: $a_{iQ}^*(a, p_i^*) = 1 \iff a \geq q_Q, a \in R$. These equilibria are all symmetric, as agents propose the same price and the same a_{iQ}^* mappings. As a consequence, full trade is attained and thus the most efficient outcome possible is reached.

In turn, when the probability of high quality λ is not large enough (namely, lower than $\lambda_1(p)$ for $p \in [q_h, r_h]$), then there is a unique symmetric equilibrium where the a_{iQ}^* mappings are the same as before but the price is the low quality cost r_l. Therefore, the adverse selection problem persists, as there exist values of λ for which only the low quality item is traded.

It is worth delving deeper into $\lambda_1(p)$, in order to clarify whether going from *homo oeconomicus* to slightly morally concerned agents significantly alleviates the inefficiencies stemming from the unequal distribution of information. The function is illustrated in Figure 3. Firstly, notice that for $p \in [q_h, r_h], \lambda_1(p) \in (0, 1)$. In addition, $\lambda_1(p)$ is increasing in p:

$$\lambda_1'(p) = \frac{v_h - r_h}{(p_j - r_l + v_h - r_h)^2} > 0.$$

Moreover (again for $p \in [q_h, r_h]$), $\lambda_1(p) \in [0, \lambda_e]$:

$$0 \leq \frac{p - r_l}{p_j - r_l + v_h - r_h} \leq \frac{r_h - r_l}{v_h - r_l}.$$

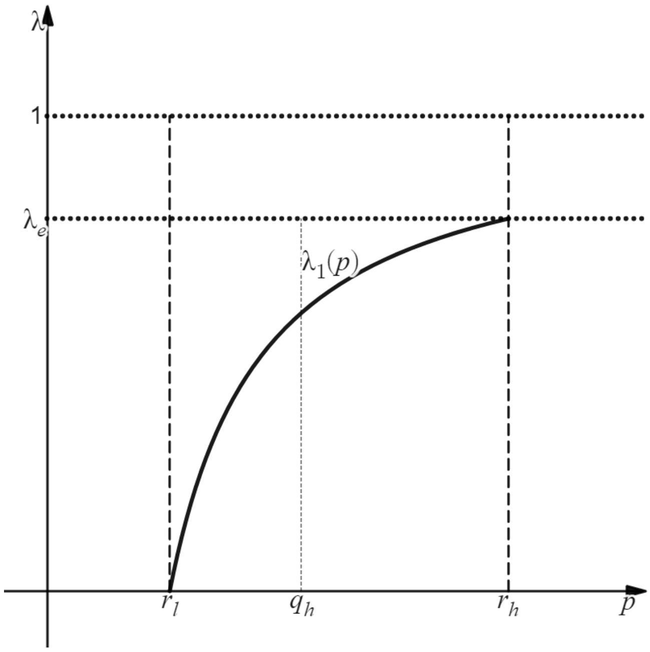


Figure 3: Function $\lambda_1(p)$

It follows that the λ interval for which moral agents are able to sustain fully efficient equilibria is larger to that of *homo oeconomicus* players. We can then conclude that a mild degree of morality (no larger than κ_1) does improve on the situation, although not in the dramatic way seen in Rivero-Wildemauwe (2023), where a tiny yet positive degree of morality sufficed to do away with non-efficient equilibria. The lowest possible value of λ that can sustain full trade equilibria corresponds to $\lambda_1(q_h)$. But notice that q_h is actually a function of κ:

$$q_h(\kappa) = \frac{r_h - \kappa v_h}{1 - \kappa}.$$

We can thus substitute $q_h(\kappa)$ in $\lambda_1(p)$ to obtain the minimal value of λ necessary to sustain a full trade equilibrium as a function of κ: $\lambda_{min} : [0, \frac{r_h - r_l}{v_h - r_l}] \to [0, 1]$. The function $\lambda_{min}(\kappa)$ is depicted in Figure 4. As the degree of morality increases and approaches κ_1, the lower bound of the price interval for which full trade takes place (q_h) falls (it equals r_l when $\kappa = \kappa_1$). In consequence, the lowest value of λ than can support a full trade equilibrium falls as well, all the way down to zero when κ reaches κ_1.

The game's equilibria when the degree of morality surpasses κ_1 but not κ_2 are laid out in Proposition 5.3. Recall that for $\kappa \in [\kappa_1, \kappa_2)$, $\mathbf{q_l} \leq \mathbf{q_h} \leq r_l < v_l < r_h < v_h$. That is, the moral reservation price for the low quality item is still below that for the high quality, but the latter is lower than both costs.

Proposition 5.3. *Consider the asymmetric information game and assume $\kappa \in [\kappa_1, \kappa_2)$. Then the game's equilibria according to Definition 4.1 are:*

- *For any λ: $p_i^* = p_j^* = p \in [q_h, r_l], a_{iQ}^*(p, p_i^*) = 1$, for all $Q \in \{h, l\}; i, j \in \{1, 2\}; i \neq j$.*
- *When $\lambda \geq \lambda_1(p)$ for every $p \in [r_l, r_h]$ and every $\lambda_1(p)$: $p_i^* = p_j^* = p \in [r_l, r_h], a_{iQ}^*(p, p_i^*) = 1$, for all $Q \in \{h, l\}$, $i, j \in \{1, 2\}$, $i \neq j$.*

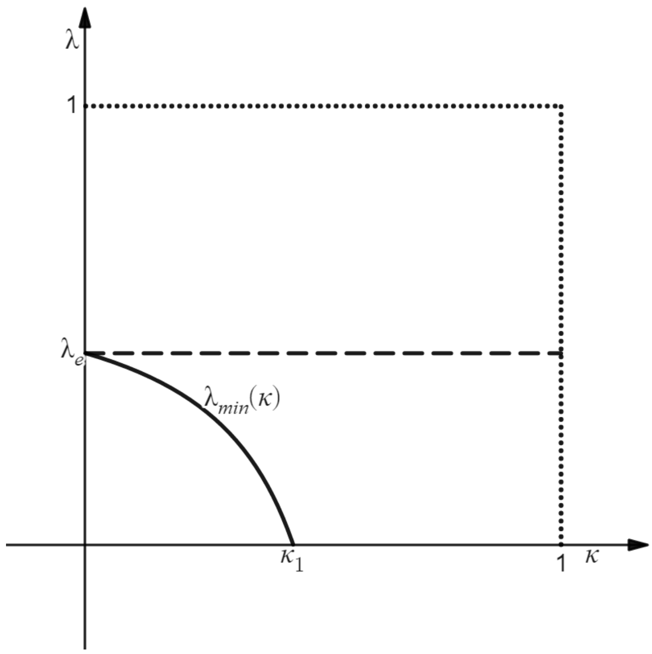


Figure 4: Function $\lambda_{min}(\kappa)$

The first noteworthy effect of agents having a larger degree of morality (at least as high as κ_1) is that there are no more inefficient equilibria. That is, only equilibria where both qualities are traded remain. In contrast with the case where $\kappa \in (0, \kappa_1)$, here when considering the function $\lambda_1(p)$, its domain is the entire $[r_l, r_h]$ interval.

This result comes about because with $\kappa \geq \kappa_1$, both moral reservation prices sit below the costs. As a consequence, the required minimum level of λ that sustains a full trade equilibrium goes as low as zero. This is easily seen in Figure 3, where the function $\lambda_1(\not{p})$ equals zero then $\not{p} = r_l$. In other words, even for $\lambda = 0$, we can find p such that $\lambda \geq \lambda_1(p)$, namely r_l. The prices that support full trade equilibria are then $p^* \in [q_h, \not{p}]$, where $\not{p} \in [r_l, r_h]$ is such that $\lambda_1(\not{p}) \leq \lambda$.

Finally, when agents exhibit a high degree of morality, with $\kappa \geq \kappa_2$, the same effect obtains. The difference is that now the lower bound on the price interval that sustains the equilibria is the moral reservation price for the low quality, q_l and not q_h. Proposition 5.4 presents this result.

Proposition 5.4. *Consider the asymmetric information game and assume $\kappa \geq \kappa_2$. Then the game's equilibria according to Definition 4.1 are:*

- *For any λ: $p_i^* = p_j^* = p \in [q_l, r_l], a_{iQ}^*(p, p_i^*) = 1$, for all $Q \in \{h, l\}; i, j \in \{1, 2\}; i \neq j$.*
- *When $\lambda \geq \lambda_1(p)$ for every $p \in [r_l, r_h]$ and every $\lambda_1(p)$: $p_i^* = p_j^* = p \in [r_l, r_h], a_{iQ}^*(p, p_i^*) = 1$, for all $Q \in \{h, l\}$, $i, j \in \{1, 2\}$, $i \neq j$.*

6 Discussion

Moral concerns have been shown to have a sound theoretical support (Alger and Weibull (2013)) and to be empirically relevant (Miettinen et al. (2020), Van Leeuwen and Alger (2023) Alger and Rivero-Wildemauwe (2023)). While efforts to understand the effects of moral preferences on different contexts have been made (see Section 1), so far the literature has not engaged with bilateral trade situations where informational asymmetries are present, except for Rivero-Wildemauwe (2023).

In this paper, I analyse the efficiency properties of a bilateral trade game's equilibria where bids are posted sequentially rather than simultaneously (as in Rivero-Wildemauwe (2023)). While a simultaneous action setting naturally does away with market power, in the present study the *Buyer* is a monopsonist, as she moves first. This set-up allows me to stay clear of potential signalling issues, as the party who holds an informational advantage is indeed the *Seller*.

I find that the expected quality required for moral agents to avoid the adverse selection outcome is weakly lower than that of *homo oeconomicus* players. Furthermore, it is decreasing in the degree of morality. As a consequence, agents with a low but strictly positive degree of morality are able to reach fully efficient equilibria for a broader range of expected quality values as compared to the model populated by pure profit maximisers. This entails that the range of prices that can support these equilibria is also enlarged, with sellers willing to accept prices lower than their cost. I also find a threshold degree of morality above which all equilibria are efficient. Considering moral agents in this setting therefore mitigates the adverse selection problem and completely eliminates it when the degree of morality is sufficiently high. These results are "smoother" as compared to Rivero-Wildemauwe (2023). With the sequential set-up used here, going from *homo oeconomicus* agents to mildly moral ones only modifies the equilibrium set marginally, as opposed to Rivero-Wildemauwe (2023) where a strictly positive degree of morality does away with any inefficient result in most situations.

A relevant remark to be made is that the experimental literature typically finds that proposers in the Ultimatum Game do not attempt to keep all the surplus generated and responders are not willing to accept an "unfair" distribution of the surplus. This fact has been explained by a mixture of fairness concerns, inequity aversion and reciprocity (Camerer (2003), Camerer and Thaler (1995), Fehr and Fischbacher (2003), Charness and Rabin (2002)) in the presence of risk aversion. Briefly put, if a responder might experience dis-utility from knowing that she has received an "unfair" offer, she might refuse. Knowing this, a risk-averse proposer may not offer a very in-egalitarian distribution. More recently, Capraro and Rodriguez-Lara (2021) report experimental results where both the amount proposed and the minimum accepted offer depend positively on proxies of moral concerns.

Why are then these results (and the model proposed in Alger and Weibull (2013)) seemingly at odds with the current findings? Central to the answer are the facts that the payoff functions used here are "linear in money" and that agents know exactly what their rival is willing to accept or not, even after factoring in moral concerns. Linearity implies that what is "moral" for these agents is to behave in a way that conduces to the surplus being produced, not caring about its distribution. In turn, with partial morality, sequentiality means that when proposing, agents choose a price that maximises their share of the surplus *subject* to being willing to accept said price if in the role of respondent. My results are thus in line with Eriksson et al. (2017), where framing the reject option as "*Payoff Reduction*" leads rejection rates to decline and the option to be judged morally wrong, as opposed to when the choice is simply labelled as "*Reject*". They are also coherent with the relatively high acceptance rate of responders in the "mini" Ultimatum Game played behind the Veil of Ignorance reported by Van Leeuwen and Alger (2023).

The current study suggests a number of avenues for future research. A natural first question to be asked is what happens when it is the *Seller* who bids first and thus holds monopoly power. The interest in providing an answer lies in the fact that the seller of a high quality item might be interested in signalling this to the *Buyer* through her proposed price (Wolinsky (1983), Jones and Hudson (1996)). Secondly, the comparison with other types of pro-social preferences such as inequity aversion (Fehr and Schmidt (1999)) in this framework is also pertinent: are the equilibria of a game between inequity-averse individuals more efficient? Are they in line with the experimental literature? Thirdly, this paper considers agents with the same degree of morality, and moreover, this is common knowledge. A follow-up question is what would happen in the case where players hold different degrees of morality. In addition, one cannot help but to wonder what would the equilibria look like if agents did not know their counterpart's degree of morality but could form an expectation about it. The issue seems to be particularly relevant in the presence of risk aversion.

Finally, this framework could serve as a template for future experimental research, as the game analysed seems to be applicable in a laboratory setting. It would be of particular interest to study individuals' behaviour in treatments where the adverse selection problem is more or less acute and have them decide behind and in front of the Veil-of-Ignorance (as in Alger and Rivero-Wildemauwe (2023)). By the same token, eliciting players' degree of morality or concerns for fairness and making it common knowledge before choices are made would provide a good test for how far players in the *Buyer* role are willing to go in their impulse to maximise their material payoff.

References

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Alger, I. and R. Renault (2007). Screening ethics when honest agents keep their word. *Economic theory*, 291–311.

Alger, I. and J. Weibull (2016). Morality: evolutionary foundations and policy implications. TSE Working Papers 16-702, Toulouse School of Economics (TSE).

Alger, I. and J. W. Weibull (2012). Homo moralis-preference evolution under incomplete information and assortative matching.

Alger, I. and J. W. Weibull (2013). Homo moralis: Preference evolution under incomplete information and assortative matching. *Econometrica 81*(6).

Alger, I. and J. W. Weibull (2017). Strategic behavior of moralists and altruists. *Games 8*(3).

Alger, I., J. W. Weibull, and L. Lehmann (2020). Evolution of preferences in structured populations: Genes, guns, and culture. *Journal of Economic Theory 185*, 104951.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal 100*(401), 464–477.

Arrow, K. J. (1973). Rawls's principle of just saving. *The Swedish Journal of Economics 75*(4), 323–335.

Ausubel, L. M., P. Cramton, and R. J. Deneckere (2002). Bargaining with incomplete information. *Handbook of game theory with economic applications 3*, 1897–1945.

Ayoubi, C. and B. Thurm (2020, August). Pro-environmental behavior and morality: An economic model with heterogeneous preferences. *SSRN*.

Ayoubi, C. and B. Thurm (2023). Knowledge diffusion and morality: Why do we freely share valuable information with strangers? *Journal of Economics & Management Strategy 32*(1), 75–99.

Becker, G. S. (1976). Altruism, egoism, and genetic fitness: Economics and sociobiology. *Journal of Economic Literature 14*(3), 817–826.

Benabou, R. and J. Tirole (2006, December). Incentives and prosocial behavior. *American Economic Review 96*(5), 1652–1678.

Bernheim, B. D. (1994). A theory of conformity. *Journal of political Economy 102*(5), 841–877.

Bilancini, E., L. Boncinelli, V. Capraro, T. Celadin, and R. Di Paolo (2020). " do the right thing" for whom? an experiment on ingroup favouritism, group assorting and moral suasion. *arXiv preprint arXiv:2002.12464*.

Bursztyn, L., S. Fiorin, D. Gottlieb, and M. Kanz. Moral incentives in credit card debt repayment: Evidence from a field experiment. *Journal of Political Economy 127*(4).

Camerer, C. and R. H. Thaler (1995). Anomalies: Ultimatums, dictators and manners. *Journal of Economic perspectives 9*(2), 209–219.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.

Capraro, V. and D. G. Rand (2018). Do the right thing: Experimental evidence that preferences for moral behavior, rather than equity or efficiency per se, drive human prosociality. *Judgment and Decision Making 13*(1), 99—-111.

Capraro, V. and I. Rodriguez-Lara (2021). Moral preferences in bargaining games. *Available at SSRN 3933603*.

Capraro, V. and A. Vanzo (2019). The power of moral words: Loaded language generates framing effects in the extreme dictator game. *Judgment and Decision Making 14*(3), 309–317.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics 117*(3), 817–869.

Coase, R. H. (1960). The problem of social cost. *The Journal of Law and Economics 3*, 1–44.

Dufwenberg, M., P. Heidhues, G. Kirchsteiger, F. Riedel, and J. Sobel (2011). Other-regarding preferences in general equilibrium. *The Review of Economic Studies 78*(2), 613–639.

Ellingsen, T. and M. Johannesson (2008, June). Pride and prejudice: The human side of incentive theory. *American Economic Review 98*(3), 990–1008.

Englmaier, F. and A. Wambach (2010). Optimal incentive contracts under inequity aversion. *Games and Economic Behavior 69*(2), 312–328.

Eriksson, K., P. Strimling, P. A. Andersson, and T. Lindholm (2017). Costly punishment in the ultimatum game evokes moral concern, in particular when framed as payoff reduction. *Journal of Experimental Social Psychology 69*, 59–64.

Fehr, E. and U. Fischbacher (2003). The nature of human altruism. *Nature 425*(6960), 785–791.

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

Harsanyi, J. C. (1953). Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy 61*(5), 434–435.

Iriberri, N. and P. Rey-Biel (2011). The role of role uncertainty in modified dictator games. *Experimental Economics 14*, 160–180.

Jones, P. and J. Hudson (1996). Signalling product quality: When is price relevant? *Journal of Economic Behavior and Organization 30*(2), 257–266.

Laslier, J.-F. (2023, May). Universalization and altruism. *Social Choice and Welfare 60*(4), 579–594.

Leland, H. E. and D. H. Pyle (1977). Informational asymmetries, financial structure, and financial intermediation. *The Journal of Finance 32*(2), 371–387.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics 1*(3), 593 – 622.

Levine, S., M. Kleiman-Weiner, L. Schulz, J. Tenenbaum, and F. Cushman (2020). The logic of universalization guides moral judgment. *Proceedings of the National Academy of Sciences 117*(42), 26158–26169.

Miettinen, T., M. Kosfeld, E. Fehr, and J. Weibull (2020). Revealed preferences in a sequential prisoners' dilemma: A horse-race between six utility functions. *Journal of Economic Behavior & Organization 173*, 1–25.

Muñoz Sobrado, E. (2022). Taxing moral agents.

Myerson, R. B. and M. A. Satterthwaite (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory 29*(2), 265 – 281.

Rawls, J. (1957). Symposium: Justice as fairness. *Journal of Philosophy 54*(22), 653–62.

Rivero-Wildemauwe, J. I. (2023). Trade among moral agents with information asymmetries. THEMA Working Papers 2023-10, THEMA (Théorie Economique, Modélisation et Applications), CY Cergy Paris Université.

Rothschild, M. and J. Stiglitz (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics 90*(4), 629–649.

Sarkisian, R. (2017). Team incentives under moral and altruistic preferences: Which team to choose? *Games 8*(3).

Stiglitz, J. E. and A. Weiss (1981). Credit rationing in markets with imperfect information. *The American Economic Review 71*(3), 393–410.

Sutter, M., J. Huber, M. Kirchler, M. Stefan, and M. Walzl (2020). Where to look for the morals in markets? *Experimental Economics 23*, 30–52.

Tappin, B. M. and V. Capraro (2018). Doing good vs. avoiding bad in prosocial choice: A refined test and extension of the morality preference hypothesis. *Journal of Experimental Social Psychology 79*, 64 – 70.

Van Leeuwen, B. and I. Alger (2023). Estimating social preferences and kantian morality in strategic interactions.

Vickrey, W. (1945). Measuring marginal utility by reactions to risk. *Econometrica: Journal of the Econometric Society*, 319–333.

Wolinsky, A. (1983). Prices as signals of product quality. *The Review of Economic Studies 50*(4), 647–658.

A Appendix

A.1 Accept-or-reject stage (asymmetric information and moral agents)

I begin by figuring out i's optimal accept or reject mappings, $a_{iQ}^{BR}(p, p_i), Q = \{h, l\}$. These are the same for all $\kappa \in (0, 1]$.

Lemma A.1. *If $p_i = p_j = p; i \neq j$, the Best Response mapping $a_{iQ}^{BR}(p, p)$ for $Q = \{h, l\}$ is*

$$a_{iQ}^{BR}(p, p) = 1 \iff p \geq q_Q \text{ and } a_{iQ}^{BR}(p, p) = 0 \text{ otherwise.}$$

If $p_i \neq p_j$, then for $Q = \{h, l\}$,

$$\begin{aligned} &a_{iQ}^{BR}(p_i, p_i) = 1 \text{ for all } p_i \neq p_j \\ &a_{iQ}^{BR}(p_j, p_i) = 1 \iff p \geq r_Q \text{ and } a_{iQ}(p_j) = 0 \text{ otherwise.} \end{aligned}$$

To expedite the reading of the proofs, it is useful to establish the following Definition:

Definition A.2. *Define the probability that the item is of quality Q as P_Q. We thus have $P_h = \lambda$ and $P_l = 1 - P_h = 1 - \lambda$.*

Proof. Consider $\cancel{a}_{iQ} = 1$ and $\cancel{a}'_{iQ} = 0$:

$$\begin{aligned}
&(1-\kappa) \cdot \Big[\lambda a_{jh}(p,p)(v_h - p) + (1-\lambda) a_{jl}(p,p)(v_l - p) + P_Q \cancel{a}_{iQ}(p - r_Q) + (1 - P_Q) a_{iR}(p,p)(p - r_R)\Big] + \\
&\kappa \cdot \Big[P_Q \cancel{a}_{iQ}(v_Q - r_Q) + (1 - P_Q) a_{iR}(p,p)(v_R - r_R)\Big] \geq \\
&(1-\kappa) \cdot \Big[\lambda a_{jh}(p,p)(v_h - p) + (1-\lambda) a_{jl}(p,p)(v_l - p) + P_Q \cancel{a}'_{iQ}(p - r_Q) + (1 - P_Q) a_{iR}(p,p)(p - r_R)\Big] + \\
&\kappa \cdot \Big[P_Q \cancel{a}'_{iQ}(v_Q - r_Q) + (1 - P_Q) a_{iR}(p,p)(v_R - r_R)\Big] \iff \\
&(1-\kappa) \cdot \Big[\lambda a_{jh}(p,p)(v_h - p) + (1-\lambda) a_{jl}(p,p)(v_l - p) + P_Q(p - r_Q) + (1 - P_Q) a_{iR}(p,p)(p - r_R)\Big] + \\
&\kappa \cdot \Big[P_Q(v_Q - r_Q) + (1 - P_Q) a_{iR}(p,p)(v_R - r_R)\Big] \geq \\
&(1-\kappa) \cdot \Big[\lambda a_{jh}(p,p)(v_h - p) + (1-\lambda) a_{jl}(p,p)(v_l - p) + (1 - P_Q) a_{iR}(p,p)(p - r_R)\Big] + \\
&\kappa \cdot \Big[(1 - P_Q) a_{ib}(p,p)(v_R - r_R)\Big] \iff \\
&p \geq \frac{r_Q - \kappa v_Q}{1 - \kappa} \equiv q_Q
\end{aligned}$$

In turn, if $p_i \neq p_j$, we need to actually distinguish $a_{iQ}(p_i, p_i)$ and $a_{iQ}(p_j, p_i)$. For $Q = \{h, l\}$, given that $(v_Q - r_Q) > 0$, it is clear that $a_{iQ}(p_i, p_i) = 1$ for all $p_i \neq p_j$ (the agent always wants to "trade with herself"). As for $a_{iQ}(p_j, p_i)$, the agent will not sell to the other player at prices lower than costs. We thus have:

$$
\begin{aligned}
&a_{iQ}(p_i, p_i) = 1 \text{ for all } p_i \neq p_j \\
&a_{iQ}(p_j, p_i) = 1 \iff p \geq r_Q \text{ and } a_{iQ}(p_j) = 0 \text{ otherwise.}
\end{aligned}
\tag{19}
$$

□

A.2 Price-setting stage (asymmetric information and moral agents)

I compute here the price best responses $p_i^{BR}(p_j)$. Recall that according to Definition 4.1, these ought to be a Best Reply to p_j *given* the a_{iQ}^{BR}, a_{jQ}^{BR} mappings (with $Q \in \{h, l\}$). For this stage, it is necessary to distinguish between different cases as κ surpasses the thresholds κ_1 and κ_2.

Lemma A.3.

For $\kappa \in (0, \kappa_2)$:

1. *If $p_j \geq r_h$: $p_i^{BR} = r_h$ when $\lambda \geq \lambda_e$, $p_i^{BR} = r_l$ otherwise*
2. *If $p_j \in [q_h, r_h)$: $p_i^{BR} = p_j$ when $\lambda \geq \lambda_1(p_j)$, $p_i = r_l$ otherwise. Notice that for $p_j \in [r_l, r_h)$, we have $\lambda_1(p_j) < \lambda_e$.*
3. *If $p_j \in [r_l, q_h)$: $p_i^{BR} = r_h$ when $\lambda \geq \lambda_e$, $p_i^{BR} = r_l$ otherwise*
4. *If $p_j \in [q_l, r_l)$: $p_i^{BR} = r_h$ when $\lambda \geq \lambda_e$, $p_i^{BR} = r_l$ otherwise*
5. *If $p_j < q_l$: $p_i^{BR} = r_h$ when $\lambda \geq \lambda_e$, $p_i^{BR} = r_l$ otherwise*

Proof. I find the price p_i that maximises (10) given a_{iQ}^{BR}, a_{jQ}^{BR} (with $Q \in \{h, l\}$), for all possible values of p_j.

$\mathbf{p_j \geq r_h}$

a. If $p_i = p_j = p \geq r_h \implies a_{iQ}^{BR}(p, p) = 1$ for all $Q = \{h, l\}, i = \{1, 2\}$.

b. If $p_i \neq p_j$, set lowest p_i possible

   b1. $p_i = r_h + \epsilon, \epsilon \to 0, \implies a_{iQ}^{BR}(p_j, p_i) = 1$ for all $Q = \{h, l\}; i, j \in \{1, 2\}$

   b2. $p_i = r_l \implies a_{iQ}^{BR}(p_j, p_i) = 1$ for all $Q = \{h, l\}$, $a_{iQ}^{BR}(p_i, p_i) = 1$ for all $Q = \{h, l\}$; $a_{jh}^{BR}(p_i, p_j) = 0$ and $a_{jl}^{BR}(p_i, p_j) = 1$

Between (a) and (b1), the only difference is that i buys at a larger price. So (b1) is always preferred to (a). **The Best Response is r_h if $\lambda \geq \lambda_e$ and r_l otherwise**.

$\mathbf{p_j \in [q_h, r_h)}$:

a. If $p_i = p_j = p \in [q_h, r_h) \implies a_{iQ}^{BR}(p, p) = 1$ for all $Q = \{h, l\}, i = \{1, 2\}$.

b. If $p_i \neq p_j$, set lowest p_i possible

b1. $p_i = r_h \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 1, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 1, a_{jl}^{BR}(p_i, p_j) = 1$

b2. $p_i = r_l \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j) = 1, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_i) = 0, a_{jl}^{BR}(p_i, p_j) = 1$

Start by comparing (a) to (b1):

$$
\begin{aligned}
&U_a\big((p_j, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \geq U_{b1}\big((r_h, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \iff \\
&(1-\kappa) \cdot \Big[\lambda(v_h - p_j) + (1-\lambda)(v_l - p_j) + \lambda(p_j - r_h) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] > \\
&(1-\kappa) \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_h) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff \\
&r_h \geq p_j \text{ (true by assumption)}
\end{aligned}
$$

Therefore, between (a) and (b1), always choose (a). Now compare (a) to (b2):

$$
\begin{aligned}
&U_a\big((p_j, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \geq U_{b2}\big((r_l, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \iff \\
&(1-\kappa) \cdot \Big[\lambda(v_h - p_j) + (1-\lambda)(v_l - p_j) + \lambda(p_j - r_h) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \geq \\
&(1-\kappa) \cdot \Big[(1-\lambda)(v_l - r_l) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff \\
&\lambda \geq \frac{p_j - r_l}{p_j - r_l + v_h - r_h} \equiv \lambda_1(p_j)
\end{aligned}
$$

Notice that $\lambda_1(p_j) \in (0,1)$. In addition, recall that we assume $\kappa < \kappa_1 \implies r_l < q_h$ and we are in the case $p_j \leq r_h$. So $p_j \in (r_l, r_h]$. Therefore, $\lambda_1(p_j) \leq \lambda_e$. The **Best Response is then $\mathbf{p_i = p_j = p \in [q_h, r_h]}$ if and only if $\lambda \geq \boldsymbol{\lambda_1(p_j)}$ and $\mathbf{r_l}$ otherwise**.

$\mathbf{p_j \in [r_l, q_h)}$:

a. If $p_i = p_j = p \in [r_l, q_h) \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 1, a_{ih}^{BR}(p_i, p_i) = 0, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 0, a_{jl}^{BR}(p_i, p_j) = 1$

b. If $p_i \neq p_j$, set lowest p_i possible

b1. $p_i = r_h \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 1, a_{ih}^{BR}(p_i, p_j) = 1, a_{il}^{BR}(p_i, p_j) = 1, a_{jh}^{BR}(p_i, p_j) = 1, a_{jl}^{BR}(p_i, p_j) = 1$

b2. $p_i = r_l \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 1, a_{ih}^{BR}(p_i, p_j) = 1, a_{il}^{BR}(p_i, p_j) = 1, a_{jh}^{BR}(p_i, p_j) = 0, a_{jl}^{BR}(p_i, p_j) = 1$

Start by comparing (a) to (b2):

$$
\begin{aligned}
&U_a\big((p_j, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \geq U_{b2}\big((r_l, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \iff \\
&(1-\kappa)\cdot\Big[(1-\lambda)(v_l - p_j) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa\cdot\Big[(1-\lambda)(v_l - r_l)\Big] \geq \\
&(1-\kappa)\cdot\Big[(1-\lambda)(v_l - r_l) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa\cdot\Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff \\
&(1-\kappa)(1-\lambda)(r_l - p_j) - \kappa\lambda(v_h - r_h) \leq 0 \text{ (impossible)}
\end{aligned}
$$

Therefore between (a) and (b2), the agent always chooses (b2). Now, compare (b2) with (b1).

$$
\begin{aligned}
&U_{b1}\big((r_h, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \geq U_{b2}\big((r_l, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \iff \\
&(1-\kappa)\cdot\Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_h) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa\cdot\Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \geq \\
&(1-\kappa)\cdot\Big[(1-\lambda)(v_l - r_l) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa\cdot\Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff \\
&\lambda \geq \lambda_e
\end{aligned}
$$

The **Best Response is then $\mathbf{r_h}$ if and only if $\lambda \geq \lambda_\mathbf{e}$ and $\mathbf{r_l}$ otherwise**.

$\mathbf{p_j} \in [\mathbf{q_l}, \mathbf{r_l})$:

a. If $p_i = p_j = p \in [q_l, r_l) \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 1, a_{ih}^{BR}(p_i, p_i) = 0, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 0, a_{jl}^{BR}(p_i, p_j) = 1$

b. If $p_i \neq p_j$, set lowest p_i possible

   b1. $p_i = r_h \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 1, a_{jl}^{BR}(p_i) = 1$

   b2. $p_i = r_l \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 0, a_{jl}^{BR}(p_i, p_j) = 1$

Start by comparing (a) to (b2):

$$U_a\big((p_j, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh^{BR}}, a_{jl}^{BR})\big) \geq U_{b2}\big((r_l, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \iff$$
$$(1-\kappa)\cdot\Big[(1-\lambda)(v_l - p_j) + (1-\lambda)(p_j - r_l)\Big]+$$
$$\kappa\cdot\Big[(1-\lambda)(v_l - r_l)\Big] \geq$$
$$(1-\kappa)\cdot\Big[(1-\lambda)(v_l - r_l)\Big]+$$
$$\kappa\cdot\Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff$$
$$-\kappa\lambda(v_h - r_h) \geq 0 \text{ (impossible)}$$

Therefore between (a) and (b2), the agent always chooses (b2). Now, compare (b2) with (b1).

$$U_{b1}\big((r_h, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \geq U_{b2}\big((r_l, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \iff$$
$$(1-\kappa)\cdot\Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_h)\Big]+$$
$$\kappa\cdot\Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \geq$$
$$(1-\kappa)\cdot\Big[(1-\lambda)(v_l - r_l) + (1-\lambda)(p_j - r_l)\Big]+$$
$$\kappa\cdot\Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff$$
$$\lambda \geq \lambda_e$$

The **Best Response is then $\mathbf{r_h}$ if and only if $\lambda \geq \lambda_\mathbf{e}$ and $\mathbf{r_l}$ otherwise**.

$\mathbf{p_j < q_l}$:

a. If $p_i = p_j = p \in [q_l, r_l) \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 0, a_{il}^{BR}(p_i, p_i) = 0, a_{jh}^{BR}(p_i, p_j) = 0, a_{jl}^{BR}(p_i, p_j) = 0$

b. If $p_i \neq p_j$, set lowest p_i possible

   b1. $p_i = r_h \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_i) = 1, a_{jl}^{BR}(p_i, p_i) = 1$

   b2. $p_i = r_l \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 0, a_{jl}^{BR}(p_i, p_j) = 1$

The **Best Response is then $\mathbf{r_h}$ if and only if $\lambda \geq \lambda_\mathbf{e}$ and $\mathbf{r_l}$ otherwise**.

□

Lemma A.4.

For $\kappa \in [\kappa_1, \kappa_2)$, we have that $q_l \leq q_h < r_l$. The price best replies are:

1. *If $p_j \geq r_h$: $p_i^{BR} = r_h$ when $\lambda \geq \lambda_e$, $p_i^{BR} = r_l$ otherwise*

2. *If $p_j \in [r_l, r_h)$: $p_i^{BR} = p_j$ when $\lambda \geq \lambda_1(p_j)$, $p_i^{BR} = r_l$ otherwise. Notice that for $p_j \in [r_l, r_h)$, we have $\lambda_1(p_j) < \lambda_e$.*

3. If $p_j \in [q_h, r_l)$: $p_i^{BR} = p_j$

4. If $p_j \in [q_l, q_h)$: $p_i^{BR} = r_h$ when $\lambda \geq \lambda_e$, $p_i^{BR} = r_l$ otherwise

5. If $p_j < q_l$: $p_i^{BR} = r_h$ when $\lambda \geq \lambda_e$, $p_i^{BR} = r_l$ otherwise

Proof. I find the price p_i that maximises (10) given a_{iQ}^{BR}, a_{jQ}^{BR} (with $Q \in \{h, l\}$), for all possible values of p_j.

$\mathbf{p_j \geq r_h}$

a. If $p_i = p_j = p \geq r_h \implies a_{iQ}^{BR}(p, p) = 1$ for all $Q = \{h, l\}, i = \{1, 2\}$.

b. If $p_i \neq p_j$, set lowest p_i possible

   b1. $p_i = r_h + \epsilon, \epsilon \to 0, \implies a_{iQ}^{BR}(p_j, p_i) = 1$ for all $Q = \{h, l\}; i, j \in \{1, 2\}$

   b2. $p_i = r_l \implies a_{iQ}^{BR}(p_j, p_i) = 1$ for all $Q = \{h, l\}$, $a_{iQ}^{BR}(p_i, p_i) = 1$ for all $Q = \{h, l\}$; $a_{jh}^{BR}(p_i, p_j) = 0$ and $a_{jl}^{BR}(p_i, p_j) = 1$

Between (a) and (b1), the only difference is that i buys at a larger price. So (b1) is always preferred to (a). **The Best Response is r_h if $\lambda \geq \lambda_e$ and r_l otherwise**.

$\mathbf{p_j \in [r_l, r_h)}$

a. If $p_i = p_j = p \in [r_l, r_h) \implies a_{iQ}^{BR}(p, p) = 1$ for all $Q = \{h, l\}, i = \{1, 2\}$.

b. If $p_i \neq p_j$, set lowest p_i possible

   b1. $p_i = r_h \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 1, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 1, a_{jl}(p_i, p_j) = 1$

   b2. $p_i = r_l \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j) = 1, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_i) = 0, a_{jl}^{BR}(p_i, p_j) = 1$

Start by comparing (a) to (b1):

$$
\begin{aligned}
& U_a\big((p_j, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) > U_{b1}\big((r_h, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \iff \\
& (1-\kappa) \cdot \Big[\lambda(v_h - p_j) + (1-\lambda)(v_l - p_j) + \lambda(p_j - r_h) + (1-\lambda)(p_j - r_l)\Big] + \\
& \kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] > \\
& (1-\kappa) \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_h) + (1-\lambda)(p_j - r_l)\Big] + \\
& \kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff \\
& r_h > p_j \text{ (true by assumption)}
\end{aligned}
$$

So between (a) and (b1), always choose (a). Now compare (a) to (b2):

$$U_a\big((p_j,a_{ih}^{BR},a_{il}^{BR}),(p_j,a_{jh}^{BR},a_{jl}^{BR})\big) > U_{b2}\big((r_l,a_{ih}^{BR},a_{il}^{BR}),(p_j,a_{jh}^{BR},a_{jl}^{BR})\big) \iff$$
$$(1-\kappa)\cdot\Big[\lambda(v_h-p_j)+(1-\lambda)(v_l-p_j)+\lambda(p_j-r_h)+(1-\lambda)(p_j-r_l)\Big]+$$
$$\kappa\cdot\Big[\lambda(v_h-r_h)+(1-\lambda)(v_l-r_l)\Big] >$$
$$(1-\kappa)\cdot\Big[(1-\lambda)(v_l-r_l)+(1-\lambda)(p_j-r_l)\Big]+$$
$$\kappa\cdot\Big[\lambda(v_h-r_h)+(1-\lambda)(v_l-r_l)\Big] \iff$$
$$\lambda > \frac{p_j-r_l}{p_j-r_l+v_h-r_h} \equiv \lambda_1(p_j)$$

Notice that $\lambda_1(p_j) \in (0,1)$ and moreover, $\lambda_1(p_j) < \lambda_e$. The **Best Response is then $\mathbf{p_i = p_j = p \in [r_l, r_h)}$ if and only if $\lambda \geq \lambda_1(\mathbf{p_j})$ and $\mathbf{r_l}$ otherwise**. Notice that $\lambda_1(r_l) = 0$ and therefore, $p_i^{BR}(r_l) = r_l$ and it induces full trade.

$\mathbf{p_j} \in [\mathbf{q_h}, \mathbf{r_l})$

a. If $p_i = p_j = p \in [q_h, r_l) \implies a_{ih}^{BR}(p_j,p_i) = 1, a_{il}^{BR}(p_j,p_i) = 1, a_{ih}^{BR}(p_i,p_i) = 1, a_{il}^{BR}(p_i,p_i) = 1, a_{jh}^{BR}(p_i,p_j) = 0, a_{jl}^{BR}(p_i,p_j) = 1$

b. If $p_i \neq p_j$, set lowest p_i possible

  b1. $p_i = r_h \implies a_{ih}^{BR}(p_j,p_i) = 0, a_{il}^{BR}(p_j,p_i) = 0, a_{ih}^{BR}(p_i,p_i) = 1, a_{il}^{BR}(p_i,p_i) = 1, a_{jh}^{BR}(p_i,p_j) = 1, a_{jl}(p_i,p_j) = 1$

  b2. $p_i = r_l \implies a_{ih}^{BR}(p_j,p_i) = 0, a_{il}^{BR}(p_j,p_i) = 0, a_{ih}^{BR}(p_i,p_i) = 1, a_{il}^{BR}(p_i,p_i) = 1, a_{jh}^{BR}(p_i,p_j) = 0, a_{jl}^{BR}(p_i,p_j) = 1$

Start by comparing (a) to (b1):

$$U_a\big((p_j,a_{ih},a_{il}),(p_j,a_{jh},a_{jl})\big) \geq U_{b1}\big((r_h,a_{ih},a_{il}),(p_j,a_{jh},a_{jl})\big) \iff$$
$$(1-\kappa)\cdot\Big[\lambda(v_h-p_j)+(1-\lambda)(v_l-p_j)+\lambda(p_j-r_h)+(1-\lambda)(p_j-r_l)\Big]+$$
$$\kappa\cdot\Big[\lambda(v_h-r_h)+(1-\lambda)(v_l-r_l)\Big] \geq$$
$$(1-\kappa)\cdot\Big[\lambda(v_h-r_h)+(1-\lambda)(v_l-r_h)\Big]+$$
$$\kappa\cdot\Big[\lambda(v_h-r_h)+(1-\lambda)(v_l-r_l)\Big] \iff$$
$$1 \geq \lambda \text{ (true by definition)}$$

So between (a) and (b1), the agent always chooses (a). Now, compare (a) with (b2).

$$
\begin{aligned}
&U_a\big((p_j, a_{ih}, a_{il}), (p_j, a_{jh}, a_{jl})\big) \geq U_{b2}\big((r_l, a_{ih}, a_{il}), (p_j, a_{jh}, a_{jl})\big) \iff \\
&(1-\kappa) \cdot \Big[\lambda(v_h - p_j) + (1-\lambda)(v_l - p_j) + \lambda(p_j - r_h) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \geq \\
&(1-\kappa) \cdot \Big[(1-\lambda)(v_l - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff \\
&\lambda(v_h - r_h) \geq 0 \text{ (true by assumption)}
\end{aligned}
$$

The **Best Response is $\mathbf{p_j}$**.

$\mathbf{p_j \in [q_l, q_h)}$

a. If $p_i = p_j \in [q_l, q_h) \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 1, a_{ih}^{BR}(p_i, p_i) = 0, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i) = 0, a_{jl}^{BR}(p_i) = 1$

b. If $p_i \neq p_j$, set lowest p_i possible

b1. $p_i = r_h \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 1, a_{jl}(p_i, p_j) = 1$

b2. $p_i = r_l \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 0, a_{jl}^{BR}(p_i, p_j) = 1$

Start by comparing (a) to (b2):

$$
\begin{aligned}
&U_a\big((p_j, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) > U_{b2}\big((r_l, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \iff \\
&(1-\kappa) \cdot \Big[(1-\lambda)(v_l - p_j) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa \cdot \Big[(1-\lambda)(v_l - r_l)\Big] > \\
&(1-\kappa) \cdot \Big[(1-\lambda)(v_l - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff \\
&-\kappa\lambda(v_h - r_h) > 0 \text{ (impossible by assumption)}
\end{aligned}
$$

So between (a) and (b2), the agent always chooses (b2). Now, compare (b2) with (b1).

$$
\begin{aligned}
&U_{b1}\big((r_h, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \geq U_{b2}\big((r_l, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \iff \\
&(1-\kappa) \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_h)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \geq \\
&(1-\kappa) \cdot \Big[(1-\lambda)(v_l - r_l) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff \\
&\lambda \geq \lambda_e
\end{aligned}
$$

The **Best Response is $\mathbf{r_h}$ if and only if $\lambda \geq \lambda_\mathbf{e}$ and $\mathbf{r_l}$ otherwise**.

$\mathbf{p_j < q_l}$

a. If $p_i = p_j = p < q_l \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 0, a_{il}^{BR}(p_i, p_i) = 0, a_{jh}^{BR}(p_i, p_j) = 0, a_{jl}^{BR}(p_i, p_j) = 0$

b. If $p_i \neq p_j$, set lowest p_i possible

b1. $p_i = r_h \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}(p_i, p_i) = 1, a_{jl}(p_i, p_i) = 1$

b2. $p_i = r_l \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 0, a_{jl}^{BR}(p_i, p_j) = 1$

The **Best Response is then $\mathbf{r_h}$ if and only if $\lambda \geq \lambda_\mathbf{e}$ and $\mathbf{r_l}$ otherwise**.

□

Lemma A.5.

For $\kappa \geq \kappa_2$, we have that $q_h \leq q_l < r_l$. The price best replies are:

1. *If $p_j \geq r_h$: $p_i^{BR} = r_h$ when $\lambda \geq \lambda_e$, $p_i^{BR} = r_l$ otherwise*
2. *If $p_j \in [r_l, r_h)$: $p_i^{BR} = p_j$ when $\lambda \geq \lambda_1(p_j)$, $p_i^{BR} = r_l$ otherwise. Notice that for $p_j \in [r_l, r_h)$, we have $\lambda_1(p_j) < \lambda_e$.*
3. *If $p_j \in [q_l, r_l)$: $p_i^{BR} = p_j$*
4. *If $p_j \in [q_h, q_l)$: $p_i^{BR} = r_h$ when $\lambda \geq \lambda_e$, $p_i^{BR} = r_l$ otherwise*
5. *If $p_j < q_l$: $p_i^{BR} = r_h$ when $\lambda \geq \lambda_e$, $p_i^{BR} = r_l$ otherwise*

Proof. I find the price p_i that maximises (10) given a_{iQ}^{BR}, a_{jQ}^{BR} (with $Q \in \{h, l\}$), for all possible values of p_j.

$\mathbf{p_j \geq r_h}$

a. If $p_i = p_j = p \geq r_h \implies a_{iQ}^{BR}(p, p) = 1$ for all $Q = \{h, l\}, i = \{1, 2\}$.

b. If $p_i \neq p_j$, set lowest p_i possible

b1. $p_i = r_h + \epsilon, \epsilon \to 0, \implies a_{iQ}^{BR}(p_j, p_i) = 1$ for all $Q = \{h, l\}; i, j \in \{1, 2\}$

b2. $p_i = r_l \implies a_{iQ}^{BR}(p_j, p_i) = 1$ for all $Q = \{h, l\}$, $a_{iQ}^{BR}(p_i, p_i) = 1$ for all $Q = \{h, l\}$; $a_{jh}^{BR}(p_i, p_j) = 0$ and $a_{jl}^{BR}(p_i, p_j) = 1$

Between (a) and (b1), the only difference is that i buys at a larger price. So (b1) is always preferred to (a). **The Best Response is r_h if $\lambda \geq \lambda_\mathbf{e}$ and r_l otherwise**.

$\mathbf{p_j \in [r_l, r_h)}$

a. If $p_i = p_j = p \in [r_l, r_h) \implies a_{iQ}^{BR}(p,p) = 1$ for all $Q = \{h, l\}, i = \{1, 2\}$.

b. If $p_i \neq p_j$, set lowest p_i possible

b1. $p_i = r_h \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 1, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 1, a_{jl}(p_i, p_j) = 1$

b2. $p_i = r_l \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j) = 1, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_i) = 0, a_{jl}^{BR}(p_i, p_j) = 1$

Start by comparing (a) to (b1):

$$
\begin{aligned}
&U_a\big((p_j, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) > U_{b1}\big((r_h, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \iff \\
&(1-\kappa) \cdot \Big[\lambda(v_h - p_j) + (1-\lambda)(v_l - p_j) + \lambda(p_j - r_h) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] > \\
&(1-\kappa) \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_h) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff \\
&r_h > p_j \text{ (true by assumption)}
\end{aligned}
$$

So between (a) and (b1), always choose (a). Now compare (a) to (b2):

$$
\begin{aligned}
&U_a\big((p_j, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) > U_{b2}\big((r_l, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \iff \\
&(1-\kappa) \cdot \Big[\lambda(v_h - p_j) + (1-\lambda)(v_l - p_j) + \lambda(p_j - r_h) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] > \\
&(1-\kappa) \cdot \Big[(1-\lambda)(v_l - r_l) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff \\
&\lambda > \frac{p_j - r_l}{p_j - r_l + v_h - r_h} \equiv \lambda_1(p_j)
\end{aligned}
$$

Notice that $\lambda_1(p_j) \in (0, 1)$ and moreover, $\lambda_1(p_j) < \lambda_e$. The **Best Response is then $\mathbf{p_i = p_j = p \in [r_l, r_h)}$ if and only if $\lambda \geq \lambda_1(\mathbf{p_j})$ and $\mathbf{r_l}$ otherwise**. Notice that $\lambda_1(r_l) = 0$ and therefore, $p_i^{BR}(r_l) = r_l$ and it induces full trade.

$\mathbf{p_j \in [q_l, r_l)}$

a. If $p_i = p_j = p \in [q_l, r_l) \implies a_{ih}^{BR}(p_j, p_i) = 1, a_{il}^{BR}(p_j, p_i) = 1, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 0, a_{jl}^{BR}(p_i, p_j) = 1$

b. If $p_i \neq p_j$, set lowest p_i possible

b1. $p_i = r_h \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 1, a_{jl}(p_i, p_j) = 1$

b2. $p_i = r_l \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 0, a_{jl}^{BR}(p_i, p_j) = 1$

Start by comparing (a) to (b1):

$$
\begin{aligned}
&U_a\big((p_j, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \geq U_{b1}\big((r_h, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \iff \\
&(1-\kappa) \cdot \Big[\lambda(v_h - p_j) + (1-\lambda)(v_l - p_j) + \lambda(p_j - r_h) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \geq \\
&(1-\kappa) \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_h)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff \\
&1 \geq \lambda \text{ (true by definition)}
\end{aligned}
$$

So between (a) and (b1), the agent always chooses (a). Now, compare (a) with (b2).

$$
\begin{aligned}
&U_a\big((p_j, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \geq U_{b2}\big((r_l, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \iff \\
&(1-\kappa) \cdot \Big[\lambda(v_h - p_j) + (1-\lambda)(v_l - p_j) + \lambda(p_j - r_h) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \geq \\
&(1-\kappa) \cdot \Big[(1-\lambda)(v_l - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff \\
&\lambda(v_h - r_h) \geq 0 \text{ (true by assumption)}
\end{aligned}
$$

The **Best Response is $\mathbf{p_j}$**.

$\mathbf{p_j \in [q_h, q_l)}$

a. If $p_i = p_j \in [q_h, q_l) \implies a_{ih}^{BR}(p_j, p_i) = 1, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 0, a_{jh}^{BR}(p_i) = 1, a_{jl}^{BR}(p_i) = 0$

b. If $p_i \neq p_j$, set lowest p_i possible

b1. $p_i = r_h \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 1, a_{jl}(p_i, p_j) = 1$

b2. $p_i = r_l \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 0, a_{jl}^{BR}(p_i, p_j) = 1$

Start by comparing (a) to (b2):

$$
\begin{aligned}
&U_a\big((p_j, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) > U_{b2}\big((r_l, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \iff \\
&(1-\kappa) \cdot \Big[(1-\lambda)(v_l - p_j) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa \cdot \Big[(1-\lambda)(v_l - r_l)\Big] > \\
&(1-\kappa) \cdot \Big[(1-\lambda)(v_l - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff \\
&-\kappa\lambda(v_h - r_h) > 0 \text{ (impossible by assumption)}
\end{aligned}
$$

So between (a) and (b2), the agent always chooses (b2). Now, compare (b2) with (b1).

$$
\begin{aligned}
&U_{b1}\big((r_h, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \geq U_{b2}\big((r_l, a_{ih}^{BR}, a_{il}^{BR}), (p_j, a_{jh}^{BR}, a_{jl}^{BR})\big) \iff \\
&(1-\kappa) \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_h)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \geq \\
&(1-\kappa) \cdot \Big[(1-\lambda)(v_l - r_l) + (1-\lambda)(p_j - r_l)\Big] + \\
&\kappa \cdot \Big[\lambda(v_h - r_h) + (1-\lambda)(v_l - r_l)\Big] \iff \\
&\lambda \geq \lambda_e
\end{aligned}
$$

The **Best Response is $\mathbf{r_h}$ if and only if $\lambda \geq \lambda_e$ and $\mathbf{r_l}$ otherwise**.

$\mathbf{p_j < q_h}$

a. If $p_i = p_j = p < q_h \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 0, a_{il}^{BR}(p_i, p_i) = 0, a_{jh}^{BR}(p_i, p_j) = 0, a_{jl}^{BR}(p_i, p_j) = 0$

b. If $p_i \neq p_j$, set lowest p_i possible

  b1. $p_i = r_h \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}(p_i, p_i) = 1, a_{jl}(p_i, p_i) = 1$

  b2. $p_i = r_l \implies a_{ih}^{BR}(p_j, p_i) = 0, a_{il}^{BR}(p_j, p_i) = 0, a_{ih}^{BR}(p_i, p_i) = 1, a_{il}^{BR}(p_i, p_i) = 1, a_{jh}^{BR}(p_i, p_j) = 0, a_{jl}^{BR}(p_i, p_j) = 1$

The **Best Response is then $\mathbf{r_h}$ if and only if $\lambda \geq \lambda_e$ and $\mathbf{r_l}$ otherwise**.

□