Dynamic Games and Applications
https://doi.org/10.1007/s13235-023-00499-7

Evolution, Investment, and Bargaining

Jack Robles[1]

Accepted: 27 February 2023

Abstract

We present an evolutionary model which allows us to study the impact relationship-specific investment has on bargaining. Agents are matched to play an investment and bargaining game. During bargaining, agents have an outside option to form a new relationship, but in exercising this option loses their current investment. We find that the stochastically stable post-investment bargaining convention is dependent on the cost of investment. In particular, the larger the cost of investment, the lower is the share of gross surplus that is received. This stands in contrast with previous studies. In addition, we find that there is under-investment. We disentangle the forces which lead to these two results.

Keywords Sunk costs · Nash demand game · Evolution · Stochastic stability · Hold-up problem

JEL Classification C78 · L14

1 Introduction

Investments are often relationship specific. A large literature has been built upon the observation that specificity of investments makes the investing party vulnerable to post-investment expropriation.[1] This vulnerability leads to under-investment unless the investing party is contractually protected. Models which use a non-cooperative game-theoretic approach support

[1] See, for example, Edlin and Reichelstein [8], Bernheim and Whinston [4], Che and Hausch [5], MacLeod and Malcomson [16], and Goldüke and Kranz [10].

✉ Jack Robles
jack.robles@vuw.ac.nz

[1] School of Economics and Finance, Victoria University, Wellington, P.O. Box 600 Wellington, New Zealand

Published online: 29 March 2023

this argument.[2] An evolutionary model, on the other hand, can give a much more optimistic result.[3] However, we argue that evolutionary models have yet to properly explore post-investment bargaining.

The current study differs from previous evolutionary models of the hold-up problem in that it focuses on post-investment bargaining. In addition, this is the first evolutionary study to explicitly model the 'story' behind the hold-up problem; the hold-up problem occurs because a relationship-specific investment cannot be used in a later relationship. We model this alternative relationship by giving agents the opportunity to rematch following a failed relationship.[4] Our model allows us to isolate the evolutionary forces on post-investment bargaining from the evolutionary forces which work purely through the investment choice. We find that the threat of hold-up reduces post-investment bargaining power. We also find under-investment. However, under-investment follows not from the post-investment bargaining, but rather from the evolutionary forces which act directly on investment.

Consider a pair of agents who are engaged in some economic activity. One (or more) of these agents makes a relationship-specific investment after which the agents bargain over the resulting surplus. Because the cost of the investment is sunk at the time of the bargaining, it has been argued that this cost should have no impact on bargaining (e.g., Klein et. al. [14]). In this case, agents bargain over the gross surplus from their relationship. The result is under-investment, because the investing party isn't properly compensated for his (marginal) investment (e.g., Grout [12], Grossman and Hart [11] and Tirole [22]).

Tröger [23] and Ellingsen and Robles [9] (TER henceforth) provide a different take. TER consider a game in which investment (by one agent) is followed by the Nash Demand game. Because the Nash Demand game is a simultaneous play game with multiple Nash equilibria, the constructed investment and bargaining game has multiple subgame perfect equilibria. TER select from these equilibria using stochastic stability (Kandori et al. [13] and Young [24, 25]). TER find that stochastic stability implies efficient investment and that the investing party captures almost all of the surplus from the relationship. In other words, TER find that evolution assigns property rights to the investing party, and she responds with efficient investment.[5]

However, TER's results say less about post-investment bargaining than one might imagine. The investing agent's ability to change her investment level makes outcomes in which she does not get a large share of the surplus unstable. The dynamic process in TER is driven by the fact that the investing party can form the belief that she will receive all the surplus following an investment level that is not currently being made. This places a lower bound on the payoff that the investing agent can receive in a stochastically stable equilibrium. The prediction in TER has more to do with the relationship between efficient investment and this lower bound and less to do with post-investment bargaining.[6]

[2] Grout [12] is the seminal example.

[3] For example, Tröger [23] and Ellingsen and Robles [9] study hold-up in an evolutionary model with one-sided investment. They find efficient investment. The mechanics behind this result are discussed below.

[4] Dawid and MacLeod [7] take a related approach. We discuss details of their approach which we feel depart from the spirit of the hold-up problem below.

[5] On the other hand, Andreozzi [1, 2] includes agents with heterogeneous investment costs in the TER model. He finds that this leads to under-investment by high cost investors.

[6] Kolm [15] assumes that agents expect a 'similar' share of the surplus following all investment levels. Some agents view 'similar' as the same absolute share of the surplus; some agents view 'similar' as meaning the same relative share of the surplus. This formulation changes the lower bound on what an agent can receive following efficient investment. However, in the end the results are no more relevant to post-investment bargaining.

We provide a model that allows us to study the impact of investment cost on post-investment bargaining. Each period, agents are matched to play a two stage investment and bargaining game. In the first stage both agents decide to invest or not, and in the second they bargain over the resulting surplus. If, in the bargaining stage, agents make compatible demands, then the game ends and agents receive their demands less their investment costs. If their demands are not compatible, then agents get nothing in the bargaining stage, but still have to pay their investment costs. However, agents who are in a failed relationship get the opportunity to form a new relationship in the following period (as in [19]). Allowing agents to rematch highlights the relationship specificity of investment; being in a failed relationship should not mean that one has missed her single opportunity, but only that her investment is wasted. We characterize the stochastically stable equilibria for our model. The stochastically stable division of post-investment surplus depends on the cost of investment. However, the share of the *gross* surplus that an agent receives is *decreasing* in her investment cost and increasing in her rival's investment cost. That is, the threat of hold-up has a doubled impact on the *net* surplus an agent receives. She is out her cost of investment as in the non-evolutionary setting, and she additionally gets a smaller share of the gross surplus. This result obtains, because 'bargaining power' in our model arises from an agent's willingness to risk being in a failed relationship in order to have a chance of getting a larger share of the pie. The larger is an agent's investment cost, the less willing she is to take such a risk. Hence, the agent with a larger investment cost gets a smaller share of the post-investment gross surplus. The stochastically stable division of surplus is contrary to the efficient allocation of property rights. However, this is not what leads to under-investment in our model. Recall that if an agent can change her investment level, then this sets a lower bound on her share of the gross surplus. It follows that when one's rival can also change his investment level, then this sets an upper bound on the share that one can receive. Under-investment occurs because the return to joint investment required to create space between these bounds is greater than the return that makes joint investment efficient. If joint investment is efficient, but has a return which is insufficient to create the required space between bounds, then the stochastically stable set includes both equilibria with efficient investment and equilibria with inefficient investment.[7]

Dawid and MacLeod [6] generalized TER to include two-sided investment. Dawid and MacLeod [6] assume that if investment is symmetric, then surplus is divided by an equal split. That is, stochastic stability is only used to determine the investment choice and bargaining after asymmetric investment. Like the current paper, Dawid and MacLeod [6] find joint investment only by assuming a large return to joint investment. Dawid and MacLeod [7] add two features to their earlier model; they assume that the investment choice is subject to error, and they include an opportunity to rematch.[8] This added structure allows Dawid and MacLeod to weaken assumptions regarding the size of gross surplus. However, like their earlier paper, this paper assumes an equal split following joint investment. Negroni and Bagnoli [17] also study two-sided investment and bargaining. As opposed to the two papers by Dawid and MacLeod, Negroni and Bagnoli allow stochastic stability to determine the post-investment division of surplus. However, because they do not include a possibility for rematching, their results do not depend on the cost of investment. Bagnoli and Negroni [3]

[7] This indeterminacy is related to that found when applying Stochastic Stability to repeated games. See Robles [20].

[8] We go into greater depth when we delve into the difference between Dawid and MacLeod [7] and the current model below.

modify their earlier model to allow asymmetric investment costs, and alternative means for dividing surplus following one-sided investment.[9]

Our results are driven by agents' ability to rematch after a failed relationship. Consequently, we should be clear why the rematching in Dawid and MacLeod [7] does not lead to similar results. In our model, a pair of agents are matched together. They then decide whether to invest or not. Following the investment choice, the agents bargain over the resulting surplus. If they do not come to an agreement on the bargaining phase, then they get an opportunity to be rematched in the following period. Dawid and MacLeod's [7] model differs in two ways. As mentioned above, in Dawid and MacLeod [7] agents bargain only when there is asymmetric investment. Hence, the question of how agents bargain following joint investment is not addressed. Instead, Dawid and MacLeod [7] are interested in factors which make joint investment more (or less) stable. To this end, Dawid and MacLeod [7] assume that errors in investment are arbitrarily more likely than mutations in bargaining demand. The consequence is to place restrictions on the beliefs that agents can form regarding what happens following a change in investment. Additionally, Dawid and MacLeod [7] allow agents to use their investment when rematched.[10] This gives an investing agent an outside option which improves her bargaining position when matched with a non-investing agent. The result is a smaller incentive to deviate from efficient investment, and a consequently weakened requirement for full investment to be stochastically stable. In the present paper, current investment has no value in any future relationship. Instead investing weakens an agent's bargaining position, because if bargaining fails, then she must make another investment in the future.

The rest of this paper is organized as follows. Section 2 presents the bargaining and investment game, the evolutionary dynamic, and the solution concepts. Section 3 presents and discusses our results. Section 4 presents variations on the model, and Sect. 5 concludes. The more technical aspects of the proofs are relegated to the "Appendix".

2 Setup

Agents α and β are matched to play a two stage game. In the first stage, α and β simultaneously decide to invest or not. An agent α (resp. β) who chooses to not invest sets $I^{\alpha} = 0$ (resp. $I^{\beta} = 0$), while if she chooses to invest she sets $I_{\alpha} = \overline{I_{\alpha}}$ (resp. $I_{\beta} = \overline{I_{\beta}}$). We write $\mathcal{I} = [I_{\alpha}, I_{\beta}]$ and use $v(\mathcal{I})$ to denote the gross value created. We use the following for specific investment pairs: $\mathcal{I}^0 = [0, 0]$, $\mathcal{I}^{\alpha} = [\overline{I_{\alpha}}, 0]$, $\mathcal{I}^{\beta} = [0, \overline{I_{\beta}}]$ and $\overline{\mathcal{I}} = [\overline{I_{\alpha}}, \overline{I_{\beta}}]$. We set $v(\mathcal{I}^0) = 0$ and define $\overline{v} \equiv v(\overline{\mathcal{I}})$, $v^{\alpha} \equiv v(\mathcal{I}^{\alpha})$, and $v^{\beta} \equiv v(\mathcal{I}^{\beta})$. We assume that investment always improves efficiency; $\overline{I_{\alpha}} < \min\{v^{\alpha}, \overline{v} - v^{\beta}\}$ and $\overline{I_{\beta}} < \min\{v^{\beta}, \overline{v} - v^{\alpha}\}$. We need further assumptions regarding the magnitude of $\overline{v}$, but we make these below to tie assumptions and results.

If neither agent invests, then the game ends with both agents getting a payoff of $u_i = 0$. On the other hand, if at least one agent invests, then the agents bargain over the resulting (gross) surplus. In particular, α demands x and β demands y. If $x + y \leq v(\mathcal{I})$, then the agents receive their demands and the game ends. In this case, $u_{\alpha} = x - I_{\alpha}$ and $u_{\beta} = y - I_{\beta}$. On the other hand, if $x + y > v(\mathcal{I})$, then the relationship fails, and neither agent receives any portion of the generated surplus. The agents enter new relationships in the following period,

[9] Under one of these alternatives, Bagnoli and Negroni observe an impact from asymmetric investment which is similar to one noted in the current study. We take note of this fact in the Conclusion after we have stated our results formally.

[10] Dawid and MacLeod model a greater specificity of investment by a greater discounting of the future.

but must make new investments for the new relationships. Both agents apply a discount factor ρ to payoffs from this new relationship.

To facilitate the use of stochastic stability, we restrict demands to a finite set. Let $D(\delta) = \{\delta, 2\delta, \dots \overline{v} - \delta\}$.[11] Following an investment pair $\mathcal{I}$, we require that demands x and y are elements of $D(\delta, v(\mathcal{I})) = [\delta, v(\mathcal{I}) - \delta] \cap D(\delta)$. The values v^{α}, v^{β}, and $\overline{v}$ are all divisible by δ. In order to avoid payoff ties, We make the genericity assumption that $D(\delta) \cap \{\overline{I_\alpha}, \overline{I_\beta}, \overline{I_\alpha} + \overline{I_\beta}\} = \emptyset$.

2.1 Evolutionary Dynamic

In each period t, there are N active α agents and N active β agents. An agent i's *characteristic* is a behavioral strategy s_i and a conditional belief σ_i. Say for the moment that i is an α agent. The strategy s_i informs us of agent i's choice following any history: $s_i(\emptyset)$ indicates the agent's investment choice, and $s_i(\mathcal{I})$ indicates her demand contingent on the investment pair I. Likewise, σ_i specifies agent i's beliefs following any history. For example, if i is an α agent, then $\sigma_i(\overline{I_\beta}|\emptyset)$ is the probability with which she believes the investment $\overline{I_\beta}$ is chosen, and $\sigma_i(y|\mathcal{I})$ is the probability of a demand y following the investment pair $\mathcal{I}$. The set of allowable beliefs is finite and includes all distributions possible given the population size.

We denote a state by θ. The period t state specifies how many agents have each possible characteristic at the end of period t. The set of possible states, Θ, is finite.

Each α (resp. β) agent active in a period is involved in exactly one match. An agent involved in a failed relationship in period $t-1$ remains active in period t. He starts period t with the same characteristic that he had at the end of period $t-1$. However, if an agent has a successful period $t-1$ relationship, then he is replaced with a child. The child inherits her parent's characteristic. In either case, the characteristic is carried from the end of period $t-1$ to the beginning of period t.

There are two means by which an agent's characteristic changes prior to playing the game in period t. Each period, each agent updates her beliefs with probability $\lambda < 1$. An updating agent observes all period $t-1$ matches, and sets her beliefs to the observed distributions. If some investment pair $\mathcal{I}$ is not made in period $t-1$, then updating leaves $\sigma_i(\cdot|\mathcal{I})$ unchanged. After updating her beliefs, an agent chooses a best (behavioral) response to her new beliefs.[12] If there is more than one best response, then each is chosen with positive probability. Following the updating draw, agents 'mutate' with probability ϵ. A mutating agent's characteristic is chosen at random based upon an exogenously given distribution which gives full support to every possible characteristic. The processes of updating and mutation combine to form a Markov chain over the state space Θ in which every transition has positive probability. Hence, an ergodic distribution, $\mu(\epsilon)$, exists.

2.2 Solution Concepts

We solve for the *stochastically stable set.* A state is stochastically stable if it is given positive probability in the limit distribution $\mu^* = \lim_{\epsilon \to 0} \mu(\epsilon)$. The calculation of the stochastically stable set is facilitated by introducing two weaker concepts: equilibrium and locally stable set. An absorbing set is a minimal set from which the population requires a mutation to

[11] The inclusion or exclusion of 0 and $\overline{v}$ from $D(\delta)$ has no impact on the final results. Proposition 2 and Lemma A.2 allow us to ignore division of surplus in which demands are not bounded away from zero.

[12] That is, even if the agent believes that the investment pair $\mathcal{I}$ has zero probability, she still sets $s_i(\mathcal{I})$ to maximize her payoff conditional on $\mathcal{I}$ occurring.

escape. We call a state an equilibrium if it is the unique element of an absorbing set. If, in state θ, every possible match yields the same outcome, then we say that θ is monomorphic.

Proposition 1 *Every absorbing set consists of a single monomorphic equilibrium.*

An equilibrium is a monomorphic state in which all agents have beliefs that discourage deviation. We note that it is possible for a non-investing agent to believe that investment leads to a negative payoff. However, the worst payoff one can expect if only the rival population invests is δ. Consequently, there are three classes of equilibria: there are non-investment equilibria in which no agents invest. There are partial investment equilibria in which agents invest in one population but not the other. In partial investment equilibria, the investing agents must get a strictly positive net payoff. Finally, there are full investment equilibria in which all agents invest. In these equilibria, every agent must get a net payoff greater than δ. Of course, in both partial investment and full investment equilibria, agent demands must add up to the total gross surplus.

Proposition 1 allows us to speak of equilibria rather than absorbing sets. Let $\overline{\Theta}$ denote the set of equilibria. Stochastically stable equilibria are found through mutation counting arguments. Let θ^1 and θ^2 denote two equilibria. We represent the transition from θ^1 to θ^2 with the *directed edge* $(\theta^1 \rightarrow \theta^2)$. Let $c(\theta^1, \theta^2)$ denote the minimum mutations required for the transition from θ^1 to θ^2. For an equilibrium θ, a collection of directed edges is a θ-tree if: (i) there is no edge departing θ, and (ii) for every equilibrium $\hat{\theta} \neq \theta$ there is a unique path of directed edges from $\hat{\theta}$ to θ. We define the cost of a tree Γ as

$$C(\Gamma) = \sum_{(\theta^1 \rightarrow \theta^2) \in \Gamma} c(\theta^1, \theta^2).$$

The stochastic potential of an equilibrium θ is $\min\{C(\Gamma) | \Gamma \text{ is an } \theta\text{-tree}\}$. The following is a restatement of Young [24, 25] Theorem 2.

Theorem 1 *An equilibrium is stochastically stable if and only if no other equilibrium has lower stochastic potential.*

The calculation of the stochastically stable set is facilitated with the use of locally stable sets [21]. A subset of $\overline{\Theta}$ is *locally stable* if it is a minimal collection of equilibria from which the population cannot escape without more than one mutation. If θ and θ' are equilibria and the population can move from θ to θ' with only a single mutation, then we write $\theta' \in M(\theta)$. The transitive closure of $M(\cdot)$ is $\overline{M}(\cdot)$. That is, if there is a sequence $\theta^1, \ldots \theta^K$ such that $\theta^{k+1} \in M(\theta^k)$, then $\theta^K \in \overline{M}(\theta^1)$. A set $\mathcal{L}$ is locally stable if $\mathcal{L} = \overline{M}(\theta)$ for every $\theta \in \mathcal{L}$. Noldeke and Samuelson [18] demonstrate that if θ is stochastically stable and $\theta' \in \overline{M}(\theta)$, then θ' is also stochastically stable. That is, stochastic stability must choose between entire locally stable sets.

For the current study, no equilibrium is, by itself, locally stable. There are, however, a number of subsets of $\overline{\Theta}$ which are locally stable. Let $z(\theta)$ denote the outcome in the equilibrium θ. If $z(\theta) = z(\theta')$, then the only difference between the two equilibria is with regards to off path beliefs. One can always move between two such equilibria through a sequence of single mutation transitions (i.e., $\theta' \in \overline{M}(\theta)$.)[13] Such transitions which change only off path beliefs are referred to as *drift*.

Given the preceding paragraph, the candidates for locally stable sets are collections of equilibria all of which have the same outcome. We will see in the sequel that some, but not

[13] As shown by Lemma A.2 in the “Appendix”.

all, such collections are locally stable. A result from Troger [23] allows us to construct trees with locally stable sets as vertices. A simplified version of Troger's result, Proposition A.1, is stated in the "Appendix".

3 Results

In this section we characterize stochastic stability. To highlight the role played by our Assumptions, we begin with an intermediate result. For both results, we state the required Assumption prior to stating the result. We then discuss the result.

Assumption 1 The return to two-sided investment is sufficiently large:
$\max\{v^\alpha, v^\beta, \frac{v^\alpha+v^\beta}{1+\rho}\} < \overline{v} - \overline{I_\alpha} - \overline{I_\beta}$.

Generally speaking, getting predictions in investment bargaining games with two-sided investment requires that the returns to joint investment be sufficiently large.[14] We will see below that to get precise prediction requires a stronger assumption.

Let $\mathcal{L}(x) = \{\theta \in \overline{\Theta} | z(\theta) = (\overline{\mathcal{I}}, x, \overline{v} - x)\}$. We see that $\mathcal{L}(x)$ is the collection of full investment equilibria in which α agents demand x, and get a net payoff of $x - \overline{I_\alpha}$. Because each $\mathcal{L}(x)$ is identified with a particular equilibrium outcome, we refer to these sets as (full investment) conventions.[15]

Proposition 2 *Let Assumption 1 hold. For δ sufficiently small, and N sufficiently large, every locally stable set is identified with a set $\mathcal{L}(x)$ for some $x \in D(\delta)$.*

An immediate conclusion from Proposition 2 is that if Assumption 2 holds, then stochastic stability implies full investment. In addition, Proposition 2 is an essential first step in characterizing the stochastically stable states. It allows us to construct trees using the sets $\mathcal{L}(x)$ as nodes. We prove Proposition 2 with standard arguments: Start with an equilibrium in which, for example, α agents are not investing. Because increased investment increases total net surplus, there are equilibrium outcomes in which only the α agents have changed their investment behavior and both populations receive a higher net payoff. Fix such an equilibrium outcome $(\mathcal{I}, x, v(\mathcal{I}) - x)$, and let β agents drift to believe that α agents will demand x following $\mathcal{I}$. Now let a single α agent mutate to invest and demand x, and let the remaining α agents update in later periods. That takes us to an equilibrium with increased investment through a sequence of single mutations transitions. It remains then only to argue that there are sets $\mathcal{L}(x)$ which cannot be escaped with a single mutation. This requires that agents in both populations receive a sufficiently large payoffs to make non-investment unattractive. Assumption 1 is designed to assure that $\overline{v}$ is large enough for this to be possible. It is easy to show that if $\max\{v^\alpha, v^\beta, \frac{v^\alpha+v^\beta}{1+\rho}\} > \overline{v} - \overline{I_\alpha} - \overline{I_\beta}$, then no convention is locally stable. Instead, every locally stable set must contain equilibria without full investment.[16]

Our main result requires a stronger Assumption.

Assumption 2 Investment costs are less than the investment complementarity:
$v^\alpha + v^\beta < \overline{v} - (\overline{I_\alpha} + \overline{I_\beta})$.

[14] See, for example, Dawid and MacLeod [6] and Negroni and Bagnoli [17].

[15] Many of the transitions that we consider require that a very small fraction of the population mutate. Absent the large population assumption in Proposition 2, many of these transition might require exactly one mutation. With the large population assumption, the number of mutations become approximately proportional to the required fraction of the population.

[16] Proposition A.3 in the "Appendix".

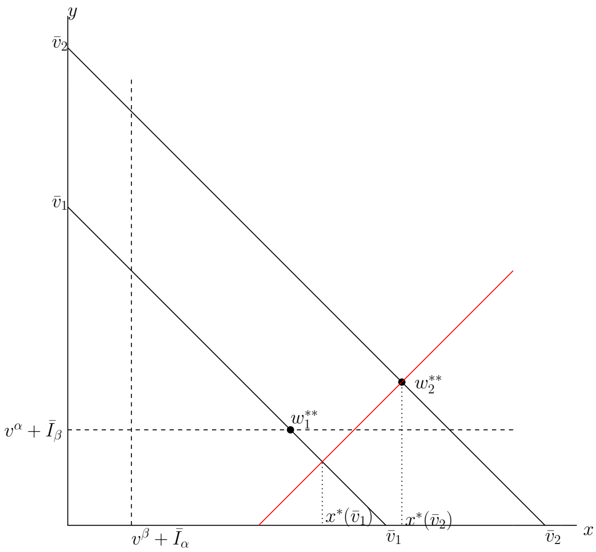

Fig. 1 Illustration of Theorem 2

Assumption 2 is a standard assumption in the literature, and is used by both Dawid and MacLeod [6] and Negroni and Bagnoli [17].

Let

$$x^* = \frac{\overline{v}}{2} + \frac{\rho(\overline{I_\beta} - \overline{I_\alpha})}{2(1-\rho)}. \tag{1}$$

The quantity x^* is central to our results.

Theorem 2 *Let Assumption 2 hold. For all δ sufficiently small and N sufficiently large, there are at most two stochastically stable conventions. The demand made by α agents in all stochastically stable conventions converges to $x^{**} = \max\{v^\beta + \overline{I_\alpha}, \min\{x^*, \overline{v} - (v^\alpha + \overline{I_\beta})\}\}$ as $\delta \to 0$.*

It is worth noting the source of the various elements of x^{**}. Recall that y denotes a β agent's demand, and let $y^* = \overline{v} - x^*$ and $y^{**} = \overline{v} - x^{**}$. The ability to change one's investment level puts a lower bound on the gross surplus received by that agent. Allocations which give that agent less are made unstable by that agent's ability to change her investment level. This is the source of the two lower bounds $x^{**} \geq v^\beta + \overline{I_\alpha}$ and $y^{**} \geq v^\alpha + \overline{I_\beta}$. Regarding x^*, imagine a model in which agents have no choice but to make a full investment. This leaves us with a model which departs [19] only in that agent's payoffs are reduced by the cost of investment. The division (x^*, y^*) would be the prediction of such a model. That is, (x^*, y^*) is the point at which the evolutionarily determined bargaining powers of the two agents are balanced.

Theorem 2 is illustrated by Figure 1 for the case $\overline{I_\alpha} > \overline{I_\beta}$.[17] In Figure 1, the lower bound on x^{**} (resp. y^{**}) is illustrated by a the dashed vertical (resp. horizontal) line. Let $\mathcal{V}(\cdot)$ denote the inverse of $x^*(\overline{v})$; $\mathcal{V}(x^*(\overline{v})) = \overline{v}$. The off diagonal red 45 degree line is the graph of $y = y^*(\mathcal{V}(x)) = x - \frac{\rho(\overline{I_\beta} - \overline{I_\alpha})}{(1-\rho)}$. The point (x^*, y^*) is found at the intersection of this line and

[17] This diagram was suggested by an anonymous referee and is a modification of a diagram found in Negroni and Bagnoli (2017).

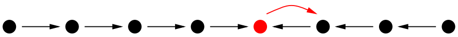

Fig. 2 Construction of a least cost tree

the gross surplus frontier $x + y = \overline{v}$. Two gross surplus frontiers are presented: $x + y = \overline{v}_1$ and $x + y = \overline{v}_2$. For $k = 1, 2$, we use $w_k^{**} = (x^{**}(\bar{v}_k), y^{**}(\bar{v}_k))$ to indicate the stochastically stable prediction. The vertical dotted lines dropping from the intersections of the red line and $x + y = \overline{v}_k$ are at $x^*(\overline{v}_k)$ $(k = 1, 2.)$ When $\overline{v} = \overline{v}_1$, we have $y^* = \overline{v}_1 - x^*(\overline{v}_1) < v^\alpha + \overline{I_\beta}$, and the stochastically stable division is at $w_1^{**} = (\overline{v}_1 - (v^\alpha + \overline{I_\beta}), v^\alpha + \overline{I_\beta})$. When $\overline{v} = \overline{v}_2$, $y^* > v^\alpha + \overline{I_\beta}$, and stochastic stability predicts the division $w_2^{**} = (x^*, y^*)$. It is apparent that as $\overline{v}$ increases, one moves from the situation in which one of the lower bounds binds and so determines the stochastically stable outcome to a situation in which neither constraint binds.

3.1 Outline of Proof

Given Proposition 2, Theorem 2 is proven by building trees with locally stable sets (full investment conventions) as nodes. Our approach parallels Young [24, 25] and Robles [19]; we show that we can construct a diagram like Fig. 2 in which each of the arrows is a least cost transition out of the departed convention.[18] Each node in Fig. 2 corresponds to a locally stable convention $\mathcal{L}(x)$. For each $\mathcal{L}(x)$, the node immediately to the right of the $\mathcal{L}(x)$ node is the node $\mathcal{L}(x+\delta)$. That is, either $\mathcal{L}(x) \to \mathcal{L}(x+\delta)$ or $\mathcal{L}(x) \to \mathcal{L}(x-\delta)$ is a least cost transition out of $\mathcal{L}(x)$. This diagram is *not* a tree, because there is no node without a departing edge. However, we are able to show that transitions $\mathcal{L}(x) \to \mathcal{L}(x+\delta)$ become more difficult as x gets larger, while transitions $\mathcal{L}(x) \to \mathcal{L}(x-\delta)$ become more difficult as x gets smaller. That is, either the red arrow or the arrow below represents the most difficult transition between conventions. For the sake of discussion, assume that the red arrow is the most difficult transition of those represented. If we delete that arrow, then we are left with a tree which must be lower cost than any other possible tree. That is, the convention represented by the red dot must be the stochastically stable convention. Because our model includes investment, there are a rich variety of transitions. Much effort is spent in the “Appendix” showing that there are two ’salient’ classes of transitions. Which of these two classes of transition is relevant depends on the value of x. The salient class of transition out of $\mathcal{L}(x)$ when $x \in [v^\beta + \overline{I_\alpha}, \overline{v} - (v^\alpha + \overline{I_\beta})]$ bears resemblance to (most of) the least cost transitions in Young [24, 25], and is used to determine x^*. We consider these transitions first.

We need some additional notation. Let

$$\eta_i(x, \mathcal{I}) = \sum_{y \le v(\mathcal{I}) - x} \sigma_i(y|\mathcal{I})$$

denote the probability that an α agent i assigns to the event that β agents demands no more than $v(\mathcal{I}) - x$ following the investment pair $\mathcal{I}$. If the α agent is part of a failed relationship, then she may have an opportunity to reset η_i in the following period. However, we assume that she takes her current value as correct, and expects it to persist. Let $\Pi_i[x, \mathcal{I}, \eta_i(x, \mathcal{I})]$ denote the discounted present value that the α agent i assigns to making a demand x following an investment pair $\mathcal{I} = (I_\alpha, I_\beta)$. Since she assigns probability $\eta_i(x, \mathcal{I})$ to her demand being

[18] In Young [24, 25] and Robles [19], the collection of least cost transition is less homogeneous.

accepted and discounts the future by ρ, $\Pi_i[x, \mathcal{I}, \eta_i(x, \mathcal{I})] = \eta_i(x, \mathcal{I}) * x - I_\alpha + \rho * [1 - \eta_i(x, \mathcal{I})] * \Pi_i[x, \mathcal{I}, \eta_i(x, \mathcal{I})]$. This solves to $\Pi_i = \frac{\eta_i(x,\mathcal{I})*x-I_\alpha}{1-\rho*[1-\eta_i(x,\mathcal{I})]}$. Likewise, let $\eta_j(y, \mathcal{I})$ denote the probability that the β agent j assigns to meeting a demand no greater than $v(\mathcal{I}) - y$ following investment pair $\mathcal{I} = (I_\alpha, I_\beta)$. The discounted present value she assigns to a demand y is $\Pi_j[y, \mathcal{I}, \eta_j(y, \mathcal{I})] = \frac{\eta_j(y,\mathcal{I})*y-I_\beta}{1-\rho[1-\eta_j(y,\mathcal{I}))]}$.

We consider a transition $\mathcal{L}(x) \to \mathcal{L}(x')$ in which: β agents mutate to demand $\overline{v} - x'$ following $\overline{\mathcal{I}}$. Following these mutations, α agents update to demand x', after which the remaining β agents update to demand $\overline{v} - x'$. At this point, the populations are in an equilibrium within $\mathcal{L}(x')$. As in Young [24, 25], this sort of transition is easiest when $x' = x - \delta$. Let r denote the portion of the β population that mutates. In order for the α population to change their demands when they update, we need

$$\Pi_i[x - \delta, \overline{\mathcal{I}}, 1] = x - \delta - \overline{I_\alpha} \geq \frac{(1-r)x - \overline{I_\alpha}}{1 - r\rho} = \Pi_i[x, \overline{\mathcal{I}}, (1-r)]. \tag{2}$$

Set $r^a(x)$ equal to the minimum value that satisfies Inequality 2. Then

$$r^a(x) = \frac{\delta}{(1-\rho)x + \rho(\delta + \overline{I_\alpha})}$$

is the minimal fraction of the β population that must mutate for the transition $\mathcal{L}(x) \to \mathcal{L}(x - \delta)$. Parallel arguments suggest that the minimum proportion of α agents that must mutate for a transition from $\mathcal{L}(x)$ to $\mathcal{L}(x + \delta)$ is

$$r^b(x) = \frac{\delta}{(1-\rho)(\overline{v} - x) + \rho(\delta + \overline{I_\beta})}.$$

As advertised, $r^a(x)$ is decreasing in x, and $r^b(x)$ is increasing in x. This leads us to define x^* by $r^a(x^*) = r^b(x^*)$. To the extent that we are focused on the correct type of transition, the convention which is most difficult to escape must be $\mathcal{L}(x)$ with $x \approx x^*$.

The reason why Theorem 2 refers to x^{**} rather that x^* is that when $x \notin [v^\beta + \overline{I_\alpha}, \overline{v} - (v^\alpha + \overline{I_\beta})]$, transitions out of $\mathcal{L}(x)$ become very easy. Consider $\mathcal{L}(x)$ with $x < v^\beta + \overline{I_\alpha} - \delta$. An α agent can increase his payoff by not investing and successfully demanding $v^\beta - \delta$. Of course, this would net the matched β agents a maximally negative payoff. If the matched β agent is not a mutant, then she would respond to such a demand by waiting in the hopes of a better match in the following period. This would shut down any transition out of $\mathcal{L}(x)$. However, if that matched β agent is also a mutant, then she might demand only δ after investing. If the other α agents observe this match, they will all imitate the α mutant and stop investing. Of course, once the β agents observe this behavior, they too will stop investing. This result is a non-investment equilibrium. Because the total surplus increases whenever investment increases, it is easy to construct a sequence of single mutation transitions that leads from this non-investment equilibrium to any full investment equilibrium. Parallel arguments apply (with roles reversed) if $x > \overline{v} - (v^\alpha + I_\beta - \delta)$. Since the number of mutations required for these transitions does not increase with population size, they become the easiest transition when N is large.

Putting these arguments together, we see that if $x^* \in [v^\beta + \overline{I_\alpha}, \overline{v} - (v^\alpha + \overline{I_\beta})]$, then the convention which is most difficult to escape is $\mathcal{L}(x)$ with $x \approx x^*$. On the other hand, if $x^* \notin [v^\beta + \overline{I_\alpha}, \overline{v} - (v^\alpha + \overline{I_\beta})]$, then it takes only two mutations to escape $\mathcal{L}(x)$ with $x \approx x^*$. In this case, the boundary element of $D(\delta) \cap [v^\beta + \overline{I_\alpha}, \overline{v} - (v^\alpha + \overline{I_\beta})]$ closest to x^* is the convention most difficult to escape. For example, if $x^* < v^\beta + \overline{I_\alpha}$, then $\forall x \in D(\delta) \cap [v^\beta + \overline{I_\alpha}, \overline{v} - (v^\alpha + \overline{I_\beta})]$

we have $r^a(x) < r^b(x)$ so that the easiest transition out of $\mathcal{L}(x)$ is to $\mathcal{L}(x-\delta)$. In this case, the most difficult convention to escape is $\mathcal{L}(x)$ with $x \approx v^\beta + \overline{I_\alpha}$ (and $x > v^\beta + \overline{I_\alpha}$). In either case, a convention is stochastically stable if and only if it is the most difficult convention to escape, and the most difficult to escape convention is $\mathcal{L}(x)$ with x within δ of x^{**}.

4 Variations on the Model

In this section, we consider some alternative formulations which are intended to illustrate our results by providing a contrast.

We begin by considering a more abstract approach. Instead of defining an agent's disagreement payoff through rematching, let us specify that, for example, an α agents expected disagreement payoff is $\mathcal{D}(I_\alpha)$. Much like a disagreement payoff which arises endogenously through rematching, the impact of this disagreement payoff is on x^* (rather than the upper and lower bounds on x^{**}). To see the import of this change, we modify an α agent i's discounted expected payoff from demand x following $\overline{\mathcal{I}}$ from Π_i to $\Pi_i^{\mathcal{D}}[x, \overline{\mathcal{I}}, \eta] = \eta x - \overline{I_\alpha} + (1-\eta)\mathcal{D}(\overline{I_\alpha})$. Assume a fraction $1-r$ of β agents demand $\overline{v} - x$ and a fraction r demand $\overline{v} - x + \delta$. The transition $\mathcal{L}(x) \to \mathcal{L}(x-\delta)$ requires that $\Pi_i^{\mathcal{D}}[x-\delta, \overline{\mathcal{I}}, 1] = x - \delta - \overline{I_\alpha} \geq (1-r)x - \overline{I_\alpha} + \mathcal{D}(\overline{I_\alpha})$. The minimal value of r which satisfies this inequality is

$$r(x|\mathcal{D}) = \frac{\delta}{x - \mathcal{D}(\overline{I_\alpha})}.$$

Assume first that $\mathcal{D}(\cdot)$ is a decreasing function; $\overline{I_\alpha} > \overline{I_\beta}$ implies $\mathcal{D}(\overline{I_\alpha}) < \mathcal{D}(\overline{I_\beta})$. In this case, $\overline{I_\alpha} > \overline{I_\beta}$ means that the α agent has an (all other things equal) inferior bargaining position. In consequence, we would see $x^* < \overline{v}/2$. We infer that this general principle is at work in our model. On the other hand, if $\mathcal{D}(\cdot)$ is an increasing function, then the agent with the larger cost of investment would have a superior bargaining position, and would receive a larger share of the pie. That is, the cost of investment impacts upon bargaining via its impact on the disagreement payoff.

We next consider a generalization motivated by comparison with Dawid and MacLeod [7].[19] Consider an agent in a failed relationship. Assume that with probability $\phi \in [0, 1]$ that agent's investment remains valid in the next match, and with probability $1-\phi$ that agent must invest anew. We must now represent an α agent i's discounted expected payoff from a demand x with a modification of Π_i which we denote Π_i^ϕ. We then have $\Pi_i^\phi[x, \overline{\mathcal{I}}, \eta] = \eta x - \overline{I_\alpha} + (1-\eta)\rho\hat{\Pi}_i^\phi$ where $\Pi_i^\phi \neq \hat{\Pi}_i^\phi$. Instead $\hat{\Pi}_i^\phi[x, \overline{\mathcal{I}}, \eta] = \eta x - (1-\phi)\overline{I_\alpha} + (1-\eta)\rho\hat{\Pi}_i^\phi$. This difference arises because a rematched agent pays $\overline{I_\alpha}$ with probability $(1-\phi)$ instead of with certainty. We note that $\hat{\Pi}_i^\phi = \frac{\eta x - (1-\phi)\overline{I_\alpha}}{1-(1-\eta)\rho}$ and that

$$\Pi_i^\phi = \hat{\Pi}_i^\phi - \phi\overline{I_\alpha} = \frac{\eta x - (1-(1-\eta)\rho\phi)\overline{I_\alpha}}{1-(1-\eta)\rho}$$

This impact of this modification is to change $r^a(x)$ and $r^b(x)$ for $x \in [v^\beta + \overline{I_\alpha}, \overline{v} - (v^\alpha + \overline{I_\beta})]$. We focus first on the impact on $r^a(x)$. Assume that a fraction r of β agents demand $\overline{v} - x + \delta$,

[19] This generalization bridge only one differences between Dawid and MacLeod [7] and the current study. Hence, it does not provide a bridge with regards to results.

while a fraction $1-r$ demand $\overline{v}-x$. The transition $\mathcal{L}(x) \to \mathcal{L}(x-\delta)$ requires

$$\Pi_i^{\phi}[x-\delta,\overline{I_\alpha},1] = x-\delta-\overline{I_\alpha} \geq \frac{(1-r)x-(1-r\rho\phi)\overline{I_\alpha}}{1-r\rho} = \Pi_i^{\phi}[x,\overline{I_\alpha},1-r].$$

The minimal value of r that satisfies this inequality is

$$r^a(x|\phi) = \frac{\delta}{(1-\rho)x+\rho(\delta+\overline{I_\alpha})-\rho\phi\overline{I_\alpha}}$$

Parallel analysis leads to

$$r^b(x|\phi) = \frac{\delta}{(1-\rho)(\overline{v}-x)+\rho(\delta+\overline{I_\beta})-\rho\phi\overline{I_\beta}}.$$

If we define x^θ by $r^a(x^\phi|\phi) = r^b(x^\phi|\phi)$, then

$$x^\phi = \frac{\overline{v}}{2} + \frac{\rho(1-\phi)(\overline{I_\beta}-\overline{I_\alpha})}{2(1-\rho)}.$$

The greater the probability $\phi > 0$ that one's investment is *not* relationship specific, the less the impact of investment cost on bargaining power. That is, it is only the relationship-specific aspect of investment that has a negative impact on bargaining power. In particular, investment made prior to bargaining is sunk; it is the potential future cost of relationship-specific investment that weakens bargaining power.

Another possible generalization is to assume that a rematch following a failed relationship occurs with probability $\xi < 1$. Again, we can define Π_i^{ξ} as a modification of Π_i which accommodates this probability. We see that $\Pi_i^{\xi}[x,\overline{\mathcal{I}},\eta] = \eta x - \overline{I_\alpha} + (1-\eta)\xi\rho\Pi_i^{\xi}$. In this case, our results would carry through but with $\xi\rho$ replacing instances of ρ.

A final modification that we might consider would be to assume that, for example, α agents choose between $\overline{I_\alpha}$ and $\underline{I_\alpha}$ with $\overline{I_\alpha} > \underline{I_\alpha} \geq 0$. This modification has no impact on x^*, because x^* is determined by bargaining power following full investment by both parties. If an α agent intends to invest $\overline{I_\alpha}$, then she will intend to again invest $\overline{I_\alpha}$ should she be rematched. That is, whether $\underline{I_\alpha} > 0$ or $\underline{I_\alpha} = 0$ is of no concern, because the α agent will not choose $\underline{I_\alpha}$. However, this modification does have an impact on x^{**}. The lower bound on x^{**} is derived from the fact that $x^{**} - \overline{I_\alpha}$ must be greater than what α can (most optimistically) hope to receive if she reduces her investment from $\overline{I_\alpha}$ to $\underline{I_\alpha}$. This hoped for payoff is now $v^\beta - \underline{I_\alpha}$ rather than v^β. Hence, the lower bound on x^{**} is $x^{**} > v^\beta + (\overline{I_\alpha} - \underline{I_\alpha})$. That is, in this alternative set up, it is the total cost of efficient investment, $\overline{I_\alpha}$, which determines post-investment bargaining power. However, it is the incremental cost of investment, $\overline{I_\alpha} - \underline{I_\alpha}$, which determines the lower bound on an α agent's share of the surplus.

5 Conclusion

We have presented an evolutionary analysis of the hold-up problem with two-sided investment. We capture the relationship specificity of investment by allowing agents to rematch after a failed relationship, but requiring them to invest anew for each new relationship. We fully characterize the stochastically stable division of surplus within our model.

Our characterization has two pieces. The first piece is the 'balance point' x^*. It is when the α agent receives x^* from the gross surplus that the evolutionarily determined bargaining strengths of the agents are balanced. The second piece is found in the lower bound on x^{**} and the symmetric lower bound on $y^{**} = \overline{v} - x^{**}$. These lower bounds arise from the ability of an investing agent to stop investing with expectation of being able to follow non-investment with a very large demand. All the studies following TER build their stochastically stable predictions from the same two pieces.[20] Our model allows us to disentangle these two pieces and the evolutionary forces that enter into them more effectively than previously.

Efficient property rights require that agents are compensated for their investment. This requires that x^{**} is increasing in $\overline{I_\alpha}$, but x^* behaves in exactly the opposite manner. That is, the evolutionarily determined bargaining power is smaller for an agent with a larger investment cost. This result obtains, because, within the dynamic process which selects x^*, disagreement in the bargaining stage is a key feature in transitions between conventions. That is, x^* is very much characterized by the degree to which agents are vulnerable to hold-up. However, while the behavior of x^* might give some insight into hold-up problems outside of the model, within our model the behavior of x^* is not the source of under-investment.

Under-investment in our model occurs because Assumption 1 is stronger than the requirement that investment is efficient. That is, it is possible for full investment to be efficient, but for outcomes with under-investment to be part of the solution. Throughout this paper, we assume that full investment is efficient: $\overline{v} - \overline{I_\alpha} > v^\beta$ and $\overline{v} - \overline{I_\beta} > v^\alpha$. For simplicity of exposition, let us assume that $\max\{v^\alpha, v^\beta, \frac{v^\alpha + v^\beta}{1+\rho}\} = v^\beta$.[21] In this case, full investment occurs when $\overline{v} - \overline{I_\alpha} - \overline{I_\beta} > v^\beta$. We notice that the LHS of this inequality is the largest payoff that an α agent could receive following full investment while still giving the β agent a nonnegative payoff. The RHS is the largest payoff that an α agent can hope to get if she does not invest. That is, the source of under-investment is an inability to satisfy the lower bound on gross payoffs implied by the ability to change one's investment level.

A key feature or our model is that we do not assume symmetry. Dawid and MacLeod [6, 7] and Negroni and Bagnoli [17] assume that $v^\alpha = v^\beta$ and $\overline{I_\alpha} = \overline{I_\beta}$. If we were to make the same assumption, we too would end up with the equal split as our prediction. Asymmetry enters our results with an impact on both x^* and the bounds on x^{**}. Because, for example, the lower bound on x^{**} is partially about avoiding the payment of $\overline{I_\alpha}$, it is the current cost of investment that plays into the bounds on x^{**}.[22] On the other hand, the impact on x^* comes from potential future investment costs. This can be seen in the fact that $x^* = \overline{v}/2$ when $\rho = 0$. On the other hand, as $\rho \to 1$, the impact of asymmetric investment costs on x^* becomes progressively more pronounced. Intuitively, if one had no investment costs, then as $\rho \to 1$ she would be willing to wait forever for a tiny increase in her payoff. However, if one does have investment costs, then he must pay these every period as he awaits that greater payoff. Consequently, as $\rho \to 1$ an agent with no investment costs has all the bargaining power even if her rival's investment costs are arbitrarily close to zero. This intuition remains valid when one agent has lower, but nonzero, investment costs.

Acknowledgements This paper has been improved by comments from participants at the 2018 Australasian Meeting of the Econometric Society, John Hillas, and anonymous referees. These comments are gratefully acknowledged.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

[20] These studies include: Andreozzi [1, 2], Dawid and MacLeod [6, 7], Kolm [15], Negroni and Bagnoli [17], and Bagnoli and Negroni [3].

[21] There is no loss in generality in assuming that $v^\beta > v^\alpha$. Ruling out $\max\{v^\alpha, v^\beta, \frac{v^\alpha + v^\beta}{1+\rho}\} = \frac{v^\alpha + v^\beta}{1+\rho}$ does come with a loss of generality. However, the arguments in this case are similar, but more technical.

[22] Bagnoli and Negroni's [3] find a similar impact from asymmetric investment costs.

Data Availability Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

A Appendix

A.1 Local Stability

In proving Proposition 2, we use a result, Proposition A.1, which is a special case of Proposition A.1 from Tröger [23]. Proposition A.1 allows us to construct trees with locally stable sets as vertices.

We extend the cost function $c(\cdot,\cdot)$ to transitions between locally stable sets. Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two locally stable sets. A *path* from $\mathcal{L}_1$ to $\mathcal{L}_2$, $P = \{(\theta_1,\theta_2),\ldots(\theta_{K-1},\theta_K)\}$, is a collection of branches between equilibrium states with $\theta_1 \in \mathcal{L}_1$ and $\theta_K \in \mathcal{L}_2$. The cost of a path P is

$$C^{-}(P) = \sum_{(\theta_k,\theta_{k+1})\in P} [c(\theta_k,\theta_{k+1}) - 1]$$

We use $\mathcal{P}(\mathcal{L}_1,\mathcal{L}_2)$ to denote the set of paths from $\mathcal{L}_1$ to $\mathcal{L}_2$. The cost of a transition from $\mathcal{L}_1$ to $\mathcal{L}_2$ is

$$c_1(\mathcal{L}_1,\mathcal{L}_2) = \min_{P\in\mathcal{P}(\mathcal{L}_1,\mathcal{L}_2)} C^{-}(P).$$

An L-tree is just like a tree, except that it has locally stable sets as vertices. The cost of an L-tree, Υ, is denote by $C_1(\Upsilon)$. We find this cost by adding up the costs of each of its branches in a manner which differs from the calculation of a normal tree only in that $c_1(\cdot,\cdot)$ is used. The L-potential of a locally stable set $\mathcal{L}$ is $\min\{C_1(\Upsilon)|\Upsilon$ is an $\mathcal{L}$-L-tree$\}$.

Proposition A.1 *A state θ is stochastically stable if and only if $\theta \in \mathcal{L}$ and $\mathcal{L}$ is a locally stable set with lowest L-potential.*

A.2 Proof of Proposition 2

Let $Z(\theta)$ denote the set of outcomes which are possible in state θ. Note, that even if the state does not change from period to period, it is possible for the outcome to change if different agents are matched together. We extend $Z(\cdot)$ to sets as follows. If $Q \subset \Theta$, then $Z(Q) = \cup_{\theta\in Q} Z(\theta)$.

Lemma A.1 *Let Q be an absorbing set. If both $(\mathcal{I}, x, y)$ and $(\mathcal{I}, x', y')$ are elements of $Z(Q)$, then $y = y' = v(\mathcal{I}) - x = v(\mathcal{I}) - x'$.*

Proof Assume otherwise. Say that $(\mathcal{I}, x, y) \in Z(\theta)$ and $\theta \in Q$. In state θ, let all α agent update and choose the same best response. In particular, there is now some $\hat{x}$ such that

$s_i(\mathcal{I}) = \hat{x}$ for every α agent i. The next time a state occurs with investment $\mathcal{I}$ made, let all β (and no α) agents update. From this point it is not possible for any demand pair other than $(\hat{x}, 1 - \hat{x})$ to occur following $\mathcal{I}$. So either the original Q was not an absorbing set or else $x = x' = \hat{x}$ and $y = y' = 1 - \hat{x}$. □

Proof of Proposition 1 We first demonstrate that absorbing sets must be singletons. After this has been established, we demonstrate that equilibria must be monomorphic. Assume Q is a non-singleton absorbing set. By Lemma A.1, each investment pair possible in Q is associated with a unique pair of demands, and so is associated with a unique payoff. By the genericity assumption, $x - \overline{I_\alpha} = x'$ (resp. $y - \overline{I_\beta} = y'$) is not possible. Hence, for any uniform investment choice by β (resp. α) agents, α (resp. β) agents must have a strict preference between investing and not investing. Hence, for Q to be a non-singleton, both α and β agents must switch between investing and not investing. Furthermore, α (resp. β) agents' preference for investing or not investing must switch with a change in β (resp. α) agents' investment choice. Finally, if β agents prefer to invest if and only if α agents invests, then α agents prefer to invest if and only if β agents don't invest.

Let us denote the four elements of $Z(Q)$ as $\mathcal{I}^0$, $(\mathcal{I}^\alpha, x^1, v^\alpha - x^1)$, $(\mathcal{I}^\beta, x^2, v^\beta - x^2)$, and $(\overline{\mathcal{I}}, x^3, \overline{v} - x^3)$. Let us assume, without loss of generality (wlog henceforth,) that the β agents prefer to invest if and only if the α agents are investing. In this case, we have that $v^\beta - x^2 - \overline{I_\beta} < 0$, and $\overline{v} - x^3 - \overline{I_\beta} > v^\alpha - x^1 \geq \delta$. Consequently, $\rho(\overline{v} - x^3 - \overline{I_\beta}) > 0 > v^\beta - x^2 - \overline{I_\beta}$.

It takes at most a sequence of 5 periods in which every α (resp. β) player gets the updating draw on odd (resp. even) periods, to reach a state θ with $Z(\theta) = \{(\overline{\mathcal{I}}, x^3, \overline{v} - x^3)\}$. From θ, let every β agent and exactly one α agent receive the updating draw. The α agent stops investing. However, whichever β agent is matched with the updating α agent expects to get $\overline{v} - x^3 - \overline{I_\beta}$ in the following period if the current match fails. Since $\rho(\overline{v} - x^3 - \overline{I_\beta}) > v^\beta - x^2 - \overline{I_\beta}$, she makes a demand greater than $v^\beta - x^2$. This results in a state, which by Lemma A.1 cannot be an element of Q. By this contradiction, an absorbing set Q must be a singleton.

We now demonstrate that equilibria are monomorphic. If $Z(\theta)$ is a non-singleton, then agents in at least one population have multiple best responses following an updating draw. This means that agents in that population change their choices with positive probability, which is to say that θ is not an equilibrium. By this contradiction, $Z(\theta)$ is a singleton. □

Since we know that all equilibria are monomorphic, it is well defined to use $z(\theta)$ to denote the unique outcome of the equilibrium θ.

Proposition A.2 *If θ is an equilibrium, then $z(\theta)$ has four possible structures:*

(1) $z(\theta) = \mathcal{I}^0$,
(2) $z(\theta) = (\mathcal{I}^\alpha, x, v^\alpha - x)$ *with* $x > \overline{I_\alpha}$,
(3) $z(\theta) = (\mathcal{I}^\beta, x, v^\beta - x)$ *with* $x < v^\beta - \overline{I_\beta}$, *and*
(4) $z(\theta) = (\overline{\mathcal{I}}, x, \overline{v} - x)$ *with* $\overline{I_\alpha} + \delta < x < \overline{v} - (\overline{I_\beta} + \delta)$.

Proof The Proposition lists all monomorphic outcomes for which it is possible to have beliefs that a deviation to a different investment level will lower an agent's payoff. Given Proposition 1, these are the possible equilibrium outcomes. □

Our next step is to identify locally stable sets.

Lemma A.2 *If θ and θ' are two equilibria with $z(\theta) = z(\theta')$, then $\theta' \in \overline{M}(\theta)$.*

Proof We note that in both θ and θ' every agent must strictly prefer to stick to the equilibrium outcome. Hence, if a mutation changes one agent's off path beliefs and actions from what they are in θ to what they are in θ', then we are still at an equilibrium. A sequence of such changes gets us to θ'. □

Let us say that $z(\theta) = (\mathcal{I}, x, v(\mathcal{I}) - x)$. If in θ all agents believe that rivals demand $v(\mathcal{I}') - \delta$ following $\mathcal{I}' \in \{\mathcal{I}^\alpha, \mathcal{I}^\beta, \overline{\mathcal{I}}\}\backslash\{\mathcal{I}\}$, then we say that agents have pessimistic beliefs. For convenience, the first step in a transition between equilibria is for agents to drift to pessimistic beliefs.

We now consider sequences of single mutation transitions that change the outcome by changing one population's investment level. For an equilibrium θ, let $u_\alpha(\theta)$ (resp. $u_\beta(\theta)$) denote the payoff that α (resp. β) agent receive in θ.

Lemma A.3 *Let $\mathcal{I} = [I_\alpha, I_\beta]$, and $\mathcal{I}' = [I'_\alpha, I'_\beta]$. Let θ and θ' be two equilibria with $z(\theta) = (\mathcal{I}, x, v(\mathcal{I}) - x)$ and $z(\theta') = (\mathcal{I}', x', v(\mathcal{I}') - x')$. The equilibrium $\theta' \in \overline{M}(\theta)$ if either of the following conditions hold:*

(1) *$I_\alpha = I'_\alpha$, $I_\beta \neq I'_\beta$, $u_\beta(\theta') > u_\beta(\theta)$ and $x' \geq \rho(x - I_\alpha) - \frac{\rho}{N}(x - x')$, or*
(2) *$I_\alpha \neq I'_\alpha$, $I_\beta = I'_\beta$, $u_\alpha(\theta') > u_\alpha(\theta)$, and $v(\mathcal{I}') - x' \geq \rho(v(\mathcal{I}) - x - I_\beta) - \frac{\rho}{N}(v(\mathcal{I}) - v(\mathcal{I}') + x' - x)$*

Proof Let $V = v([I_\alpha, I_\beta])$ and $V' = v([I'_\alpha, I'_\beta])$. The two cases are perfectly symmetric, so we only check case (1). Starting from θ, let all agent drift to pessimistic beliefs. From this new equilibrium, let all α agents drift to expect demands of $V' - x'$ following $[I_\alpha, I'_\beta]$. From this new equilibrium, let a single β agent mutate to make investment I'_β and demand $V' - x'$ following $\mathcal{I}' = [I_\alpha, I'_\beta]$. If $x' < \rho(x - I_\alpha)$, the α agent matched with the mutant will demand more than x' following $\mathcal{I}'$ in this first period. However, in the following period allow all α agents (and no β agents) to update. This sets their beliefs so that they expect (prior to seeing β's investment) that insisting on a payoff of $x - I_\alpha$ yields an expected payoff of $\Pi = \frac{(N-1)x - N \cdot I}{N-\rho}$. Upon seeing an investment of I'_β, they will be willing to demand x' if $x' - I_\alpha \geq -I_\alpha + \rho\Pi$. Canceling the $-I_\alpha$ and multiplying through by Π's denominator takes this to $(N - \rho)x' \geq \rho N(x - I_\alpha) - \rho x$ which goes to $x' \geq \rho(x - I_\alpha) - \frac{\rho}{N}(x - x')$. Hence, the α agent matched with the mutant in this period will demand x'. In the next period, let every β agent receive the updating draw. They all switch to invest I'_β and demand $V' - x'$ following $[I'_\alpha, I'_\beta]$. At this point, we are at an equilibrium which differs from θ' only for off path beliefs. Drift takes the population to θ'. □

With regard to Lemma A.3, we note that, e.g., $u_\alpha(\theta') \geq u_\alpha(\theta)$ is stronger than $x' \geq \rho(x - I_\alpha) - \frac{\rho}{N}(x - x')$. Generally, if the population changing investment strictly prefers the new equilibrium, and the population NOT changing investment weakly prefers the new equilibrium, then Lemma A.3 applies.

At this point, it is convenient to define expression $\underline{X}^e_\delta$ and $\overline{X}^e_\delta$ such that if $\underline{X}^e_\delta \leq x \leq \overline{X}^e_\delta$, then it is possible to have an equilibrium θ with $z(\theta) = (\overline{\mathcal{I}}, x, \overline{v} - x)$. From Proposition A.2, we have

$$\underline{X}^e_\delta = \min\{x \in D(\delta) | x > \overline{I_\alpha} + \delta\}, \tag{A.1}$$

$$\overline{X}^e_\delta = \max\{x \in D(\delta) | x < \overline{v} - \overline{I_\beta} - \delta\}. \tag{A.2}$$

Lemma A.4 *Fix $x \in D(\delta)$ with $\underline{X}^e_\delta \leq x \leq \overline{X}^e_\delta$. If θ^0 is a non-investment equilibrium, and $\theta \in \mathcal{L}(x)$, then $\theta \in \overline{M}(\theta^0)$.*

Proof Fix θ and x as defined in the Lemma. Let $y = \overline{v} - x$ denote the β demand in θ. Starting from a non-investment equilibrium θ^0 let all agents drift to pessimistic beliefs. Let $\hat{x} = \max\{\delta, v^\beta - y\}$. Note that $\hat{y} \equiv v^\beta - \hat{x} > \overline{I_\beta}$. Let α agents drift to expect a demand of $\hat{y}$ following $\mathcal{I}^\beta$. Let a single β agent mutate to invest and demand $\hat{y}$ following $\mathcal{I}^\beta$. In the following period, let all β agents update. This puts us at an equilibrium $\hat{\theta}$ with outcome $(\mathcal{I}^\beta, \hat{x}, v^\beta - \hat{x})$. Now α (resp. β) agents must strictly (resp. weakly) prefer an equilibrium $\theta \in \mathcal{L}(x)$ to $\hat{\theta}$. By Lemma A.3, this implies that $\theta \in \overline{M}(\hat{\theta}) \subset \overline{M}(\theta^0)$. □

Lemma A.5 *Every locally stable set contains a full investment convention $\mathcal{L}(x)$.*

Proof By Lemma A.2, the Lemma is true if one can reach a full investment equilibrium from any other equilibrium through a sequence of single mutation transitions. Given Lemma A.4, it remains only to show that this holds for partial investment equilibria.

Let θ be a partial investment equilibrium. Assume, wlog, that it is the α agents who invest and $z(\theta) = (\mathcal{I}^\alpha, x, v^\alpha - x)$. Now consider an equilibrium θ' with $z(\theta') = (\overline{\mathcal{I}}, x, \overline{v} - x)$. Since $\overline{v} > v^\alpha + \overline{I_\beta}$, β (resp. α) agents get a strictly (resp. weakly) higher payoff in θ'. Hence, Lemma A.3 applies, and $\theta' \in \overline{M}(\theta)$. □

We turn to the determination of locally stable sets. To this end, define

$$\underline{J}_\delta(x) = \{x' \in D(\delta) | x' > x - \overline{I_\alpha} \text{ and } v^\beta - x' \geq \rho(\overline{v} - x - \overline{I_\beta}) - \frac{\rho}{N}(\overline{v} - v^\beta + x' - x)\}$$

$$\underline{X}^1_\delta = \min\{x \in D(\delta) | \underline{J}_\delta(x) = \emptyset\}$$

$$\overline{J}_\delta(x) = \{x' \in D(x) | v^\alpha - x' > \overline{v} - x - \overline{I_\beta} \text{ and } x' \geq \rho(x - \overline{I_\alpha}) - \frac{\rho}{N}(x - x')\}$$

$$\overline{X}^1_\delta = \max\{x \in D(\delta) | \overline{J}_\delta(x) = \emptyset\}$$

We note that $\underline{X}^1_\delta$ and $\overline{X}^1_\delta$ are defined exactly so that we can apply Lemma A.3 to escape $\mathcal{L}(x)$ if and only if $x < \underline{X}^1_\delta$ or $x > \overline{X}^1_\delta$. Of course, if, for example, $\underline{X}^1_\delta < x < \underline{X}^e_\delta$, then $\mathcal{L}(x)$ is not well defined since there is no equilibrium with outcome $(\overline{\mathcal{I}}, x, \overline{v} - x)$. Hence, the following are used to characterize locally stable sets:

$$\underline{X}^L_\delta = \max\{\underline{X}^e_\delta, \underline{X}^1_\delta\} \tag{A.3}$$

$$\overline{X}^L_\delta = \min\{\overline{X}^e_\delta, \overline{X}^1_\delta\} \tag{A.4}$$

Lemma A.6 *If $x \in D(\delta)$ and $\underline{X}^L_\delta \leq x \leq \overline{X}^L_\delta$, then $\mathcal{L}(x)$ is a locally stable set.*

Proof Given Lemma A.2, it remains only to demonstrate that one mutation is not sufficient to escape from $\mathcal{L}(x)$. Since $\underline{X}^1_\delta \leq \underline{X}^L_\delta \leq x \leq \overline{X}^L_\delta \leq \overline{X}^1_\delta$, we know that if $x' > x - \overline{I_\alpha}$, it follows that $v^\beta - x' < \rho(\overline{v} - x' - \overline{I_\beta}) - \frac{\rho}{N}(\overline{v} - v^\beta + x' - x)$. Hence, if a single α agent were to mutate to not invest and demand x' he would not get his demand. Further, no amount of β updating will change this fact, neither will other α agents imitate him. Hence, he will eventually update which moves the population back to an element of $\mathcal{L}(x)$. Of course, if an α agent were to mutate to not invest and demand $x' < x - \overline{I_\alpha}$, then he might get his demand met. However, his payoff would be less than that of the investing α agents who would again not imitate him. Parallel arguments show that a mutation which causes a single β agent to stop investing are not sufficient to escape $\mathcal{L}(x)$. Let us now consider mutations which leave the level of investment unchanged. Say that some α agents mutate to demand x' following $\overline{\mathcal{I}}$. The attractiveness of the available options to a β agent must depend on the proportion of

α agents making different demands. For N sufficiently large, this will involve more than one mutation. □

Lemma A.6 relies upon the definitions of $\underline{X}_\delta^L$ and $\overline{X}_\delta^L$ which are difficult to use. However, our concern is only with the case of N large and δ small. By definition, $x > \overline{X}_\delta^1$ if $\exists x'$ such that $v^\alpha - x' > \overline{v} - x - \overline{I_\alpha}$ and $x' \geq \rho(x - \overline{I_\alpha}) - \frac{\rho}{N}(x - x')$. If we let $N \to \infty$, then we can rewrite this as $\exists x'$ such that $v^\alpha - (\overline{v} - x - \overline{I_\beta}) > x' \geq \rho(x - \overline{I_\alpha})$. If we consider the limit when $\delta \to 0$, then we can omit the x' safe in the knowledge that any gap between the LHS and RHS will contain elements of $D(\delta)$. Setting $u_\alpha = x - \overline{I_\alpha}$ and $u_\beta = \overline{v} - x - \overline{I_\beta}$, we have $\rho u_\alpha + u_\beta < v^\alpha$. Likewise, if we start from $x < \underline{X}_\delta^1$, parallel steps take us to $u_\alpha + \rho u_\beta < v^\beta$. We find $\underline{x}^1$ (resp. $\overline{x}^1$) below by solving $u_\alpha + \rho u_\beta = v^\beta$ (resp. $\rho u_\alpha + u_\beta = v^\alpha$).

$$\underline{x}^1 = \frac{v^\beta - \rho\overline{v} + \overline{I_\alpha} + \rho\overline{I_\beta}}{1-\rho} \tag{A.5}$$

$$\overline{x}^1 = \frac{\overline{v} - v^\alpha - \overline{I_\beta} - \rho\overline{I_\alpha}}{1-\rho} \tag{A.6}$$

Lemma A.7 (1) *As $\delta \to 0$ and $N \to \infty$, $\underline{X}_\delta^1 \to \max\{\underline{x}^1, 0\}$, and $\overline{X}_\delta^1 \to \min\{\overline{x}^1, \overline{v}\}$.*

(2) *Let Assumption 1 hold. If N is sufficiently large, and δ is sufficiently small, then $\underline{X}_\delta^L < \overline{X}_\delta^L$.*

Proof Part (1) follows from the construction of $\underline{x}^1$ and $\overline{x}^1$. Part (2): given Part (1), it suffices to show that $\max\{\underline{x}^1, \underline{X}_\delta^e\} < \min\{\overline{x}^1, \overline{X}_\delta^e\}$. Since we already knows that $\underline{X}_\delta^e < \overline{X}_\delta^e$, we must show that: (i) $\underline{x}^1 < \overline{x}^1$, (ii) $\underline{x}^1 < \overline{X}_\delta^e$, and (iii) $\underline{X}_\delta^e < \overline{x}^2$. (i): $\underline{x}^1 < \overline{x}^1 \Leftrightarrow v^\beta - \rho\overline{v} + \overline{I_\alpha} + \rho\overline{I_\beta} < \overline{v} - v^\alpha - \overline{I_\beta} - \rho\overline{I_\alpha} \Leftrightarrow \frac{v^\alpha + v^\beta}{1+\rho} < \overline{v} - \overline{I_\alpha} - \overline{I_\beta}$. (ii): we note that $\overline{X}_\delta^e \to \overline{v} - \overline{I_\beta}$. Hence, we need to show that $v^\beta - \rho\overline{v} + \overline{I_\alpha} + \rho\overline{I_\beta} < (1-\rho)(\overline{v} - \overline{I_\beta}) \Leftrightarrow v^\beta < \overline{v} - \overline{I_\alpha} - \overline{I_\beta}$. (iii): we observe that $\underline{X}_\delta^e \to \overline{I_\alpha}$. Hence, we need to show that $(1-\rho)\overline{I_\alpha} < \overline{v} - v^\alpha - \overline{I_\beta} - \rho\overline{I_\alpha} \Leftrightarrow v^\alpha < \overline{v} - \overline{I_\alpha} - \overline{I_\beta}$. In all three cases, Assumption 1 gives us what we need. □

Proof of Proposition 2 Lemma A.6 assures that $\mathcal{L}(x)$ is locally stable so long as $x \in D(\delta)$ and $\underline{X}_\delta^L < x < \overline{X}_\delta^L$. Given δ sufficiently small, and N sufficiently large, Lemma A.7 assures that such $\mathcal{L}(x)$ exist. Lemma A.5 assures that no other locally stable set exists.

Proposition A.3 *Assume that* $\max\{v^\alpha, v^\beta, \frac{v^\alpha+v^\beta}{1+\rho}\} > \overline{v} - \overline{I_\alpha} - \overline{I_\beta}$. *If δ is sufficiently small, and N is sufficiently large, then every locally stable set contains a partial investment equilibrium.*

Proof All statements are made assuming δ sufficiently small and N sufficiently large. It is immediate that $\overline{x}^1 \geq \overline{I_\alpha} \Leftrightarrow v^\alpha \leq \overline{v} - \overline{I_\alpha} - \overline{I_\beta}$, and that $\underline{x}^1 \leq \overline{v} - \overline{I_\beta} \Leftrightarrow v^\beta \geq \overline{v} - \overline{I_\alpha} - \overline{I_\beta}$. Hence, if $\max\{v^\alpha, v^\beta\} > \overline{v} - \overline{I_\alpha} - \overline{I_\beta}$, then either $\underline{x}^1 > \overline{X}_\delta^e$ or $\overline{x}^1 < \underline{X}_\delta^e$. On the other hand, $\underline{x}^1 \leq \overline{x}^1$ only if $\frac{v^\alpha+v^\beta}{1+\rho} \leq \overline{v} - \overline{I_\alpha} - \overline{I_\beta}$. Hence, if the assumption holds, then $\underline{X}_\delta^L > \overline{X}_\delta^L$. By Lemma A.3 any locally stable set which contains a full investment convention also includes a partial investment equilibrium. Since Lemma A.5 states that every locally stable set must contain a full investment convention, we are done. □

A.3 Proof of Theorem 2

We introduce two new boundaries

$$\underline{X_\delta} = \min\{x \in D(\delta) | x > v^\beta + \overline{I_\alpha} - \delta\} \tag{A.7}$$

$$\overline{X_\delta} = \max\{x \in D(\delta) | x < \overline{v} - (v^\alpha + \overline{I_\beta} - \delta)\} \tag{A.8}$$

Our remaining analysis depends heavily on whether $\underline{X_\delta} < x < \overline{X_\delta}$ or not. We first establish the relationship between $\underline{X_\delta}$ and $\overline{X_\delta}$ on the one hand, and $\underline{X_\delta}^L$ and $\overline{X_\delta}^L$ on the other.

Lemma A.8 *Let Assumption 2 hold. If δ is sufficiently small, and N is sufficiently large, then $\underline{X_\delta}^L < \underline{X_\delta} < \overline{X_\delta} < \overline{X_\delta}^L$.*

Proof That $\underline{X_\delta} < \overline{X_\delta}$ is simply a restatement of Assumption 2. For N sufficiently large, and δ sufficiently small, one may show that $\underline{X_\delta}^L < \underline{X_\delta}$ by showing that $v^\beta + \overline{I_\alpha} > \underline{x}^1 = \frac{v^\beta - \rho\overline{v} + \overline{I_\alpha} + \rho\overline{I_\beta}}{1-\rho}$. This reduces to $\overline{v} > v^\beta + \overline{I_\alpha} + \overline{I_\beta}$ which holds under Assumption 2. Likewise, to show $\overline{X_\delta} < \overline{X_\delta}^L$, it suffices to show that $\overline{v} - v^\alpha - \overline{I_\beta} > \overline{x}^1 = \frac{\overline{v} - v^\alpha - \overline{I_\alpha} - \rho\overline{I_\alpha}}{1-\rho}$ which reduces to $\overline{v} > v^\alpha + \overline{I_\alpha} + \overline{I_\beta}$. Again, this holds under Assumption 2. □

There are two types of transitions between locally stable sets $\mathcal{L}(x)$ and $\mathcal{L}(x')$. A *direct transition* from $\mathcal{L}(x)$ to $\mathcal{L}(x')$ does not pass through an equilibrium which is not an element of $\mathcal{L}(x) \cup \mathcal{L}(x')$. An *indirect transition* does pass through such a transitional equilibrium. We first establish the ease of indirect transitions out of $\mathcal{L}(x)$ when $x < \underline{X_\delta}$ or $x > \overline{X_\delta}$.

Lemma A.9 *Assume that $\mathcal{L}(x)$ is locally stable. If either $x < v^\beta + \overline{I_\alpha} - \delta$ or $x > \overline{v} - (v^\alpha + \overline{I_\beta} - \delta)$, then it takes two mutations to escape $\mathcal{L}(x)$. From that point, a sequence of single mutation transitions suffice to move the population to any alternative locally stable set.*

Proof The two cases are perfectly symmetric, so we focus only on the case when $\underline{X_\delta}^L \leq x < v^\beta + \overline{I_\beta} - \delta$. Fix an equilibrium θ with $z(\theta) = (\overline{\mathcal{I}}, x, \overline{v} - x)$ and $\underline{X_\delta}^L \leq x < v^\beta + \overline{I_\beta} - \delta$. Starting from θ, let both populations drift to pessimistic beliefs. From this equilibrium, let two mutations occur with one in each population. Let the mutating β agent choose to continue to invest, but to demand only δ after $\mathcal{I}^\beta$. Let the α mutant choose to not invest and to demand $v^\beta - \delta$ following $\mathcal{I}^\beta$. Let these two mutants be matched to play each other. The following period, let every α agent, but no β agents, update. All α agents switch to imitate the mutant. The following period, let all β but no α agents update. Since investing now nets a β agent a negative payoff, they all switch to not investing. At this point, agents in both populations expect a negative payoff from investing, which puts us at a non-investment equilibrium. Let x' be as defined in the Lemma. By Lemma A.4, we can reach $\theta' \in \mathcal{L}(x')$ from this non-investment equilibrium through a sequence of single mutations transitions. Hence, θ' can be reached from θ by a sequence of transitions one of which requires two mutations, and the rest of which require one mutation.

Lemma A.10 *Assume that $\underline{X_\delta}^L \leq x \leq \overline{X_\delta}^L$.*

If $x < \underline{X_\delta}$ and $\underline{X_\delta} < x' \leq \overline{X}_\delta^L$, then $c_1(\mathcal{L}(x), \mathcal{L}(x')) = 1$.
If $\overline{X_\delta} < x$ and $\underline{X}_\delta^L \leq x' < \overline{X_\delta}$, then $c_1(\mathcal{L}(x), \mathcal{L}(x')) = 1$.

Proof Since $\mathcal{L}(x)$ is locally stable, the transition in question must have at least two mutations at the start. With this observation, the Lemma follows from Lemma A.9. □

We turn to transition out of $\mathcal{L}(x)$ when $v^\beta + \overline{I_\alpha} - \delta < x < \overline{v} - (v^\alpha + \overline{I_\beta} - \delta)$. In this case, the number of mutations required to escape from $\mathcal{L}(x)$ is (approximately) proportional to the population size. In particular, we repeatedly find some r which is the minimal fraction of a population that must mutate for a transition. Let $\lceil \cdot \rceil^+$ denote a function which takes a real number to the smallest integer greater than or equal to that real number. A transition which requires a fraction r of a population to mutate requires $\lceil N * r \rceil^+$ mutations.

We first consider *pull* transition, in which one population 'pulls' the other to a new post-investment demand. In particular: (i) mutations change demands in one population, (ii) the second population updates to the complementary demand (i.e., $x + y = \overline{v}$,) (iii) updating leads to a new equilibrium. If (iii) lead to an equilibrium that is not part of a locally stable convention, then we add (iv) a sequence of single mutation transitions lead to a new locally stable convention. Lemma A.11 assures that we do not need to consider exotic choices for mutations at step (i) in a pull transition.

Lemma A.11 *Assume that $\underline{X}_\delta < x < \overline{X}_\delta$ and $\underline{X}_\delta^L < \hat{x} < \overline{X}_\delta^L$. Consider a pull transition away from $\mathcal{L}(x)$. If the first step is achieved through mutations to β (resp. α) agents which pull α (resp. β) agents to demand $\hat{x}$ (resp. $\overline{v} - \hat{x}$) following $\overline{I_\alpha}$, then it suffices to consider only mutations which cause β (resp. α) agents to demand $\overline{v} - \hat{x}$ (resp. $\hat{x}$) following $\overline{\mathcal{I}}$.*

Proof The roles of α and β agents are perfectly symmetric, so we prove the result only for β mutants. The expected payoff for investing and demanding x is $\Pi_i[x, \overline{\mathcal{I}}, \eta_i(x, \overline{\mathcal{I}})] = \frac{\eta_i(x,\overline{\mathcal{I}})*x-\overline{I_\alpha}}{1-\rho[1-\eta_i(x,\overline{\mathcal{I}})]}$. The expected payoff for investing and demanding $\hat{x}$ is $\Pi_i[\hat{x}, \overline{\mathcal{I}}, \eta_i(\hat{x}, \overline{\mathcal{I}})] = \frac{\eta_i(\hat{x},\overline{\mathcal{I}})*\hat{x}-\overline{I_\alpha}}{1-\rho[1-\eta_i(\hat{x},\overline{\mathcal{I}})]}$. Clearly mutations need to decrease $\Pi_i[x, \overline{\mathcal{I}}, \eta_i(x, \overline{\mathcal{I}})]$ or increase $\Pi_i[\hat{x}, \overline{\mathcal{I}}, \eta_i(\hat{x}, \overline{\mathcal{I}})]$, which can only be accomplished by decreasing $\eta_i(x, \overline{\mathcal{I}})$ or increasing $\eta_i(\hat{x}, \overline{\mathcal{I}})$. We first establish that the result holds when considering mutations which leave investment unchanged. We then rule out the possibility of mutations which change the investment choice.

If $\hat{x} > x$, then a mutation to demand $\overline{v} - x'$ increases $\eta_i(\hat{x}, \overline{\mathcal{I}})$ and (weakly) decreases $\eta_i(x, \overline{\mathcal{I}})$ if and only if $x' \geq \hat{x}$. A mutation to demand $\overline{v} - x''$ with $\hat{x} > x'' > x$ changes neither $\eta_i(x, \overline{\mathcal{I}})$ nor $\eta_i(\hat{x}, \overline{\mathcal{I}})$. A mutation to demand $\overline{v} - x''$ with $x > x''$ either leaves $\Pi_i[\hat{x}, \overline{\mathcal{I}}, \eta_i(\hat{x}, \overline{\mathcal{I}})] = 0$, or decreases both $\eta_i(x, \overline{\mathcal{I}})$ and $\eta_i(\hat{x}, \overline{\mathcal{I}})$, by the same amount. This is shown less effective than a mutation to demand $\overline{v} - x'$ with $x' \geq \hat{x}$ as follows. Let r be the fraction of β agents not demanding $\overline{v} - x$, and let w be the fraction of β agents demanding strictly more than $\overline{v} - x$. Then r and w need to satisfy

$$\frac{(r - w)\hat{x} - \overline{I_\alpha}}{1 - \rho(1 - r + w)} = \frac{(1 - w)x - \overline{I_\alpha}}{1 - \rho w} \tag{A.9}$$

Simplifying and taking a total differential of this equation yields $\frac{dr}{dw} = \frac{(\hat{x}-x)[1-\rho w+\rho(r-w)]}{\hat{x}(1-\rho w)-\rho(1-w)x+\rho I} > 0$. Hence the fraction of (fully investing) β agents not demanding $\overline{v} - x$ is larger when $w \neq 0$. Hence, when $\hat{x} > x$, mutations to demand $\overline{v} - x'$ with $x' \geq \hat{x}$ are most effective.

If $\hat{x} < x$, then a mutant demand of $\overline{v} - x'$ following $\overline{\mathcal{I}}$ (weakly) increases $\eta_i(\hat{x}, \overline{\mathcal{I}})$ and decreases $\eta_i(x, \overline{\mathcal{I}})$ if and only if $\hat{x} \leq x' < x$. A mutation to demand $\overline{v} - x''$ with $x'' > x$ changes neither $\eta_i(x, \overline{\mathcal{I}})$ nor $\eta_i(\hat{x}, \overline{\mathcal{I}})$. A mutation to demand $\overline{v} - x''$ with $x'' < \hat{x}$ decreases both $\eta_i(x, \overline{\mathcal{I}})$ and $\eta_i(\hat{x}, \overline{\mathcal{I}})$ both by the same amount. However, such a mutation decreases $\Pi_i(x, \overline{\mathcal{I}}, \eta_i(x, \overline{\mathcal{I}}))$ by the same amount as the mutation to demand $\overline{v} - x'$ with $\hat{x} \leq x' < x$. Hence, when $\hat{x} < x$, mutations to demand $\overline{v} - x'$ with $\hat{x} \leq x' < x$ are most effective. Further, as long as $x' \geq \hat{x} > x$ or $x > x' \geq \hat{x}$, the exact value of x' does not change the impact upon $\eta_i(\hat{x}, \overline{\mathcal{I}})$ or $\eta_i(x, \overline{\mathcal{I}})$. By considering mutations to demands of $\overline{v} - \hat{x}$, we remove any impact upon $\eta_i(x', \overline{\mathcal{I}})$ with $x' > \hat{x}$. This leaves demanding $\hat{x}$ as the most attractive alternative to x following $\overline{\mathcal{I}}$.

Mutations that cause agents to stop investing are demonstrated less effective than mutations to demand $\overline{v} - \hat{x}$ as follows. We first assume, as we can from above, that the only mutations which change demand but not investment lead to demands of $\overline{v} - \hat{x}$. Consider $M+1$ mutations, M of which are already specified. What form should the final mutation take? Consider first the case in which $x > \hat{x}$. There is some $K \geq N - M$ such that prior to the final mutation, $\eta_i(x, \overline{\mathcal{I}}) = (N-M)/K$. With $x > \hat{x}$, it follows that at all stages of the transition $\eta_i(\hat{x}, \overline{\mathcal{I}}) = 1$. A mutation which changes demand to $\overline{v} - \hat{x}$ changes $\eta_i(x, \overline{\mathcal{I}})$ to $\frac{M-1}{K}$. A mutation which causes β agent to stop investing changes $\eta_i(x, \overline{\mathcal{I}})$ to $\frac{M-1}{K-1}$. Since $\frac{M-1}{K} < \frac{M-1}{K-1}$ mutations to stop investing are less effective. Now consider $x < \hat{x}$. In this case, none of the mutations change the fact that $\eta_i(x, \overline{\mathcal{I}}) = 1$. Prior to the final mutation, there are $\hat{M} \leq M$ and $K \geq N - M$ such that $\eta_i(\hat{x}, \overline{\mathcal{I}}) = \hat{M}/K$. A mutation which changes demand to $\overline{v} - \hat{x}$ sets $\eta_i(\hat{x}, \overline{\mathcal{I}}) = \frac{\hat{M}+1}{K}$. A mutation which causes β agent to stop investing changes $\eta_i(\hat{x}, \overline{\mathcal{I}})$ to $\frac{\hat{M}}{K-1}$. Since $\frac{\hat{M}+1}{K} > \frac{\hat{M}}{K-1}$, mutations to stop investing are again less effective. □

Consider a pull transition out of $\theta \in \mathcal{L}(x)$, in which a fraction r of β agents mutate to demand $\overline{v} - x'$. If $x' < x$, then a transition requires

$$\Pi_i[x', \overline{\mathcal{I}}, 1] = x' - \overline{I_\alpha} \geq \frac{(1-r)x - \overline{I_\alpha}}{1 - r\rho} = \Pi_i[x, \overline{\mathcal{I}}, (1-r)]. \tag{A.10}$$

If instead $x' > x$ then a transition requires

$$\Pi_i[x', \overline{\mathcal{I}}, r] = \frac{r * x' - \overline{I_\alpha}}{1 - \rho(1-r)} \geq x - \overline{I_\alpha} = \Pi_i[x, \overline{\mathcal{I}}, 1]. \tag{A.11}$$

In both cases, a larger x' implies a smaller r. Since we are looking for least cost means for escaping $\mathcal{L}(x)$ we consider only $x' = \overline{v} - \delta$ and $x' = x - \delta$. Let $r^a(x)$ be the smallest value of r satisfying Inequality A.10 when $x' = x - \delta$. Solving yields

$$r^a(x) = \frac{\delta}{(1-\rho)x + \rho(\delta + \overline{I_\alpha})}.$$

Let $\overline{r}^a(x)$ be the smallest value of r satisfying Inequality A.11 when $x' = \overline{v} - \delta$. Solving yields

$$\overline{r}^a(x) = \frac{(1-\rho)(x - \overline{I_\alpha}) + \overline{I_\alpha}}{(\overline{v} - \delta) - \rho(x - \overline{I_\alpha})}.$$

If $v^\beta + \overline{I_\alpha} - \delta < x < \overline{v} - (v^\alpha + \overline{I_\beta}) + \delta$, then $\mathcal{L}(x - \delta)$ is a convention, and $\mathcal{L}(x) \to \mathcal{L}(x - \delta)$ is a direct transition. However, because $\overline{v} - \delta > \overline{v} - \overline{I_\beta}$, θ with $z(\theta) = (\overline{\mathcal{I}}, \overline{v} - \delta, \delta)$ is not an equilibrium. Instead, pulling α agents to demand $\overline{v} - \delta$ is a possible start of an indirect transition.

Perfectly symmetric arguments lead us to the expressions

$$r^b(x) = \frac{\delta}{(1-\rho)(\overline{v} - x) + \rho(\delta + \overline{I_\beta})}$$

and

$$\overline{r}^b(x) = \frac{(1-\rho)(\overline{v} - x - \overline{I_\alpha}) + \overline{I_\alpha}}{(\overline{v} - \delta) - \rho(\overline{v} - x - \overline{I_\alpha})}.$$

By arguments symmetric to those above, $r^b(x)$ is the fraction of the α population that must mutate for the transition $(\mathcal{L}(x) \to \mathcal{L}(x+\delta))$. Similarly, if a fraction $\overline{r}^b(x)$ of α agents mutate to demand δ following $\overline{I_\alpha}$, then that is the first step of an indirect transition.

We observe that a direct transition must be a pull transition, and that $\overline{r}^a(x) > \overline{I_\alpha}/\overline{v}$ and $\overline{r}^b(x) > \overline{I_\beta}/\overline{v}$. With these observations, Lemma A.11 and the algebra above leads to Lemma A.12.

Lemma A.12 *Assume that $\underline{X_\delta} < x < \overline{X_\delta}$. For δ sufficiently small, and N sufficiently large, the lowest cost direct transition out of $\mathcal{L}(x)$ is to either $\mathcal{L}(x-\delta)$ or to $\mathcal{L}(x+\delta)$.*

The first of these transitions takes $\lceil N * r^a(x)\rceil^+$ mutations, and the second takes $\lceil N * r^b(x)\rceil^+$.

We turn now to indirect transitions out of $\mathcal{L}(x)$. When we calculated $\overline{r}^a(x)$ above, we started with mutations to β agents which caused them to demand δ following $\overline{\mathcal{I}}$. With sufficient such mutants, α agents switched to demand $\overline{v}-\delta$ following $\overline{\mathcal{I}}$. This made investment unattractive for β agents. The same can be achieved if α agents mutate to demand $\overline{v}-\delta$ following $\overline{\mathcal{I}}$. Say that a proportion r of the α agents mutate to demand $\overline{v}-\delta$ following $\overline{\mathcal{I}}$. For β agents to think that non-investing looks better than investing we need $(1-r)(\overline{v}-x-\overline{I_\beta}) = v^\alpha - \delta$. This solves to $r = 1 - \frac{v^\alpha-\delta}{\overline{v}-x-\overline{I_\beta}}$. Likewise, we would need a proportion of β agents equal to $r = 1 - \frac{v^\beta-\delta}{x-\overline{I_\alpha}}$ to mutate to make it attractive for α agents to stop investing. Hence, an indirect transition out of $\mathcal{L}(x)$ can be achieved if a proportion

$$r^0(x) = 1 - \max\left\{\frac{v^\beta-\delta}{x-\overline{I_\alpha}}, \frac{v^\alpha-\delta}{\overline{v}-x-\overline{I_\beta}}\right\} \tag{A.12}$$

of the population mutates.

Lemma A.13 *Fix $x \in D(\delta)$ with $\underline{X_\delta} < x < \overline{X_\delta}$. Let $\Xi \subset \overline{\Theta}$ denote the set of equilibrium states without full investment.*

(i) $\min\{c(\theta,\theta')|\theta \in \mathcal{L}(x)$ *and* $\theta' \in \Xi\} = \lceil N * \min\{r^0(x), \overline{r}^a(x), \overline{r}^b(x)\}\rceil^+$.
(ii) *This lower bound can always be achieved with θ' a non-investment equilibrium.*

Proof α and β agents are symmetric, so we focus on means to get α agents to stop investing. The easiest transition out of $\mathcal{L}(x)$ must involve: α agents with the most positive possible belief regarding not investing, and (a portion) of β agents who are making the worst possible demand (from an α agent's perspective). We start by allowing drift to pessimistic beliefs. We follow this with drift so that α agents expect β agents to demand $y = \delta$ following $\mathcal{I}^\beta$. That addresses the beliefs. Regarding β demands, we want (a portion of) β agents to demand $\overline{v}-\delta$ following $\overline{\mathcal{I}}$. There are two means for achieving this. Method 1: β agents mutate to demand $\overline{v}-\delta$ following $\overline{\mathcal{I}}$. Method 2, α agents mutate to demand δ which causes updating β agents to demand $\overline{v}-\delta$ following $\overline{\mathcal{I}}$. From the algebra preceding Lemma A.12, it requires $\lceil N * \overline{r}^b(x)\rceil^+$ α mutants demanding δ following $\overline{\mathcal{I}}$ to make updating β demand $\overline{v}-\delta$ following $\overline{\mathcal{I}}$. On the other hand, if we are following Method 1, then to make α agent stop investing we need a proportion r of the β agent to mutate to solve $(1-r)(x-\overline{I_\alpha}) = v^\beta - \delta$. This solves to $r = 1 - \frac{v^\beta-\delta}{x-\overline{I_\alpha}}$. Whichever route we took to get β agents to demand $\overline{v}-\delta$, we let all α agents update. They all stop investing and demand $x = v^\beta - \delta$ following $\mathcal{I}^\beta$. In the next period, let all β agents update. They all stop investing. At this point all agents expect a negative payoff for investing, which is to say that we are at a non-investment equilibrium. Hence, if our approach is to make α agents stop investing, then it takes $\lceil N * \min\{\overline{r}^b(x), 1 - \frac{v^\beta - \delta}{x - \overline{I_\alpha}}\}\rceil^+$ mutations to achieve the transition. Parallel arguments show that if we reverse the role of the α and β agents, then we need $\lceil N * \min\{\overline{r}^a(x), 1 - \frac{v^\alpha - \delta}{\overline{v} - x - \overline{I_\beta}}\}\rceil^+$ mutations. The proof is completed by recalling the definition of $r^0(x)$ and recognizing that we will put α and β in the role that makes the transition take the fewest mutations. □

Let $D^L(\delta) = D(\delta) \cap [\underline{X}^L_\delta, \overline{X}^L_\delta]$ and $D^*(\delta) = D(\delta) \cap [v^\beta + \overline{I_\alpha} - \delta, \overline{v} - (v^\beta + \overline{I_\beta}) + \delta] = D(\delta) \cap [\underline{X}_\delta, \overline{X}_\delta]$.

Lemma A.14 *Assume that $x \in D^*(\delta)$ and $x' \in D^L(\delta)$. If the lowest cost transition from $\mathcal{L}(x)$ to $\mathcal{L}(x')$ is indirect, then for δ sufficiently small and N sufficiently large $c_1(\mathcal{L}(x), \mathcal{L}(x')) = \lceil N * \min\{r^0(x), \overline{r}^a(x), \overline{r}^b(x)\}\rceil^+ - 1$.*

Proof By Lemma A.13: we can get from $\mathcal{L}(x)$ to a non-investment equilibrium with $\lceil N * \min\{r^0(x), \overline{r}^a(x), \overline{r}^b(x)\}\rceil^+$ mutations, and fewer mutations are insufficient to move to an equilibrium without full investment. From Lemma A.4 a sequence of single mutation transition suffices to get from a non-investment equilibrium to $\mathcal{L}(x')$. That that $c_1(\mathcal{L}(x), \mathcal{L}(x')) = \lceil N * \min\{r^0(x), \overline{r}^a(x), \overline{r}^b(x)\}\rceil^+ - 1$ now follows from the definition of $c_1(\cdot, \cdot)$. □

Let $r_\delta(x) \equiv \min\{r^a(x), r^b(x), r^0(x)\}$.

Lemma A.15 *Consider $x \in D^*(\delta)$. The following statements hold for N sufficiently large and δ sufficiently small.*

(1) $\min\{c_1(\mathcal{L}(x), \mathcal{L}(x'))|x' \in D^L(\delta)\} = \lceil N * r_\delta(x)\rceil^+ - 1$
(2) *If $r^0(x) = r_\delta(x)$, then $c_1(\mathcal{L}(x), \mathcal{L}(x')) = \lceil N * r^0(x)\rceil^+ - 1$ for all $x' \in D^L(\delta)$.*
(3) *If $r^a(x) < r^0(x)$ and $x' \in D^L(\delta)$ with $x' < x - \delta$, then $c_1(\mathcal{L}(x), \mathcal{L}(x')) > c_1(\mathcal{L}(x), \mathcal{L}(x - \delta)) = \lceil N * r^a\rceil^+ - 1$.*
(4) *If $r^b(x) < r^0(x)$ and $x' \in D^L(\delta)$ with $x' > x + \delta$, then $c_1(\mathcal{L}(x), \mathcal{L}(x')) > c_1(\mathcal{L}(x), \mathcal{L}(x + \delta)) = \lceil N * r^b\rceil^+ - 1$.*

Proof (1) Follows immediately from Lemmas A.12 and A.14. (2) follows from (1) and the fact that $r^a(x) < \bar{r}^a(x)$ and $r^b(x) < \bar{r}^b(x)$. (3) and (4) follow from Lemma A.12 and the algebra which precedes it. □

One notices that for a fixed x and δ sufficiently small, $\min\{r^a(x), r^b(x)\} < r^0(x)$. We must nonetheless include $r^0(x)$ in the formula for $r_\delta(x)$ because as δ becomes smaller, $D(\delta)$ comes to include values which are closer to the boundaries $v^\beta + \overline{I_\alpha}$ and $\overline{v} - (v^\alpha + \overline{I_\beta})$.

Recall that $D^*(\delta) = D(\delta) \cap [v^\beta + \overline{I_\alpha} + \delta, \overline{v} - (v^\alpha + \overline{I_\beta}) - \delta]$. There are some immediate conclusions from thinking about $r^a(x)$, $r^b(x)$, $r^0(x)$, and $r_\delta(x)$ as functions on the interval $[v^\beta + \overline{I_\alpha} + \delta, \overline{v} - (v^\alpha + \overline{I_\beta}) - \delta]$. Since $r_\delta(x)$ is the minimum of four strictly monotonic functions, $r_\delta(x)$ is strictly quasi-concave. This leads immediately to the following.

Lemma A.16 *Let Assumption 2 hold. For δ sufficiently small, and N sufficiently large, there are at most two elements of $D^*(\delta)$ which maximize $r_\delta(x)$. If there are two such elements, then they are adjacent.*

Let us denote the elements of $D^*(\delta)$ which maximize $r_\delta(x)$ as $X^*_\delta = \{x \in D^*(\delta)|r_\delta(x) \geq r_\delta(x')\, \forall x' \in D^*(\delta)\}$. Let $\underline{x}^*_\delta$ be the smallest element of X^*_δ and let $\overline{x}^*_\delta$ be the largest. The monotonic natures of $r^a(x)$ and $r^b(x)$ assure that if $x < \overline{x}^*_\delta$, then $r^a(x) > \min\{r^b(x), r^0(x)\}$. Likewise, if $x > \underline{x}^*_\delta$, then $r^b(x) > \min\{r^a(x), r^0(x)\}$. This leads immediately to the following.

Lemma A.17 *Let Assumption 2 hold, and assume that δ is sufficiently small and that N is sufficiently large. If $x < \overline{x}^*_\delta$, then $(\mathcal{L}(x) \to \mathcal{L}(x+\delta))$ is a least cost transition out of $\mathcal{L}(x)$. If $x > \underline{x}^*_\delta$, then $(\mathcal{L}(x) \to \mathcal{L}(x-\delta))$ is a least cost transition out of $\mathcal{L}(x)$.*

Proof This is immediate from Lemma A.15 Part (3) and (4) and the monotonicity of $r^a(x)$ and $r^b(x)$. □

We don't claim that the identified transition is the only least cost transition. In particular, if $r_\delta(x) = r^0(x)$, then every transition out of $\mathcal{L}(x)$ is equally easy. The same is true if $x < v^\beta + \overline{I_\alpha} - \delta$ or $x > \overline{v} - v^\alpha - \overline{I_\beta} + \delta$. Nonetheless, it is easy to go from the results that we have to a statement that $\cup_{x \in X^*_\delta} \mathcal{L}(x)$ is the stochastically stable set. In particular, we start with a directed graph for which the nodes are the locally stable $\mathcal{L}(x)$, and each node has a single departing edge corresponding to the transitions identified in Lemma A.17. This is not a tree, but by deleting an edge with largest cost, a tree is created with the minimal possible total cost. Since the largest cost edges are those departing elements of X^*_δ, we are left with the following.

Lemma A.18 *Let Assumption 2 hold. For δ sufficiently small and N sufficiently large, $\theta \in \mathcal{L}(x)$ is stochastically stable if and only if $x \in X^*_\delta$.*

All that is required to get from Lemma A.18 to Theorem 2 is to argue that any element of X^*_δ is within δ of x^{**}. To this end, we note a couple of facts. The first fact is that x^* is defined by $r^a(x^*) = r^b(x^*)$. The second fact to note is that because $r^a(\cdot)$ and $r^b(\cdot)$ both have numerators of δ, it is only possible for $r_\delta(x) = r^0(x)$ if $x \approx v^\beta + \overline{I_\alpha}$ or $x \approx \overline{v} - (v^\alpha + \overline{I_\beta})$. (Recall that $r^0 = 1 - \max\left\{\frac{v^\beta - \delta}{x - \overline{I_\alpha}}, \frac{v^\alpha - \delta}{\overline{v} - x - \overline{I_\beta}}\right\}$). Further, the smaller is δ, the tighter need be this approximation. That is to say, it is only on the (shrinking) edges of $[v^\beta + \overline{I_\alpha} - \delta, \overline{v} - v^\alpha - \overline{I_\beta} + \delta]$ where it is possible for $r_\delta(x) = r^0(x)$. Hence, if $v^\beta + \overline{I_\alpha} < x^* < \overline{v} - (v^\alpha + \overline{I_\beta})$; then $\exists \overline{\delta} > 0$ such that if $\delta < \overline{\delta}$, then $r_\delta(\cdot)$ is maximized at x^*. In this case, it is immediate that elements of X^*_δ are within δ of x^*. On the other hand, if, for example, $x^* < v^\beta + \overline{I_\alpha}$, then we know that for δ sufficiently small $r^a(x) < r^b(x)$ for all $x \in [v^\beta + \overline{I_\alpha} - \delta, \overline{v} - v^\alpha - \overline{I_\beta} + \delta]$. This leaves two possibilities. Either $r_\delta(x)$ is maximized where $r^a(x) = r^0(x)$, or $r_\delta(x)$ is maximized where $r^a(x) < r^0(x)$. In the second case, $r_\delta(x)$ is maximized at the boundary where $x = v^\beta + \overline{I_\alpha} - \delta$. In the first case, we can see that the solution to

$$r^0(x) = \frac{x - \overline{I_\alpha} - v^\beta - \delta}{x - \overline{I_\alpha}} = \frac{\delta}{(1-\rho)(x) + \rho(\delta + \overline{I_\alpha})} = r^a(x)$$

must converge to $x = v^\beta + \overline{I_\alpha}$ as $\delta \to 0$. Hence, if $x^* \le v^\beta + \overline{I_\alpha}$, then the elements of X^*_δ must converge to $v^\beta + \overline{I_\alpha}$ as $\delta \to 0$. Parallel arguments show that if $x^* > \overline{v} - (v^\alpha + \overline{I_\beta})$, then the elements of X^*_δ must converge to $\overline{v} - (v^\alpha + \overline{I_\beta})$ as $\delta \to 0$. With these observations, Theorem 2 is proved.

References

1. Andreozzi L (2010) An evolutionary theory of social justice: choosing the right game. Eur J Polit Econ 26:320–329
2. Andreozzi L (2012) Property rights and investments: an evolutionary approach. Games Econ Behav 74:1–11
3. Bagnoli L, Negroni G (2018) Egalitarianism: an evolutionary perspective. Metroeconomica 70:24–44
4. Bernheim D, Whinston M (1998) Incomplete contracts and strategic ambiguity. Am Econ Rev 88:902–32

5. Che Y-K, Hausch D (1999) Cooperative investment and the value of contracting: Coase vs. Williamson. Am Econ Rev 89:125–47
6. Dawid H, MacLeod WB (2001) Holdup and the evolution of bargaining conventions. Eur J Econ Soc Syst 15:153–169
7. Dawid H, MacLeod WB (2008) Hold-up and the evolution of bargaining norms. Games Econ Behav 62:26–52
8. Edlin A, Reichelstein S (1996) Holdups, standard breach remedies, and optimal investment. Am Econ Rev 86:478–501
9. Ellingsen T, Robles J (2002) Does evolution solve the hold-up problem? Games Econ Behav 39:28–35
10. Goldüke S, Kranz S (2020) Reconciliating relational contracting and hold-up: a model of repeated negotiations. University of Konstanz, Konstanz
11. Grossman SJ, Hart OD (1986) The costs and benefits of ownership: a theory of of vertical and lateral integration. J Polit Econ 94:691–719
12. Grout P (1984) Investment and wages in the absence of a binding contract. Econometrica 52:449–460
13. Kandori M, Mailath G, Rob R (1993) Learning, mutation, and long run equilibria in games. Econometrica 61:29–56
14. Klein B, Crawford RG, Alchian AA (1978) Vertical integration, appropriable rents, and the competitive contracting process. J Law Econ 21:297–326
15. Kolm Julian (2011) Learning across subgames: an application to the hold-up problem. Vienna Graduate School of Economics, Vienna
16. MacLeod WB, Malcomson J (1993) Investments, holdup and the form of market contracts. Am Econ Rev 83:811–37
17. Negroni G, Bagnoli L (2017) On the coevolution of social norms in primitive societies. J Econ Interact Coord 12:635–667
18. Noldeke G, Samuelson L (1993) An evolutionary analysis of backwards and forward induction. Games Econ Behav 5:425–454
19. Robles J (2008) Evolution, bargaining, and time preferences. Econ Theor 35:19–36
20. Robles J (2011) Stochastic stability in finitely repeated two player games. BE J Theor Econ 11:767–786
21. Samuelson L (1994) Stochastic stability in games with alternative best replies. J Econ Theory 64:35–65
22. Tirole J (1986) Procurement and renegotiation. J Polit Econ 94:25–259
23. Tröger T (2002) Why sunk costs matter for bargaining outcomes: an evolutionary approach. J Econ Theory 102:375–402
24. Young P (1993) The evolution of conventions. Econometrica 61:57–84
25. Young P (1993) An evolutionary model of bargaining. J Econ Theory 59:145–168