University of Essex

Department of Economics

Discussion Paper Series

Losing Face

Thomas Gall and David Reinstein

Losing Face

Thomas Gall and David Reinstein

July 20, 2015

Abstract

When Al makes an offer to Betty that Betty *observes and* rejects, Al may "lose face". This *loss of face* (LoF) may cost Al utility, either directly or through reputation effects. This can lead to fewer offers and inefficiency in the context of bilateral matching problems, e.g., the marriage market, research partnering, and international negotiations. We offer a simple model with asymmetric information, a continuous signal of an individual's binary type, and a linear marriage production function. We add a primitive LoF term, characterize the stable equilibria, compare the benchmark without LoF to a case where only one side is vulnerable to LoF, and present comparative statics. A small amount of LoF has no effect on low types' behavior, but, will make high types on both sides more selective. A stronger LoF drives high types out of the market, and makes low types *reverse snobs*, further reducing welfare. LoF also makes rejecting *strictly* preferred to being rejected, making the "high types reject" equilibrium stable. We can eliminate the effects of LoF by letting the vulnerable side move second, or setting up a "Conditionally Anonymous Environment" that only reveals when both parties say yes. We motivate our model with a variety of empirical examples, and we suggest policy and managerial implications.

1 Introduction

In a market that involves two-sided matching (as surveyed in Burdett and Coles, 1999), the fear of rejection can lead to inefficiency. A proposer may not ask someone out on a date, ask for a study partner, apply for a job, make a business proposition, propose a paper co-authorship, or suggest a peace treaty, because she does not want the other party to *know* of her interest and then turn her down; this may have consequences for reputation and future play, or it may have a direct psychological cost. In general, we call the disutility from this potential outcome "loss of face" (LoF).

Consider a game where each player can choose "accept" or "reject" and there is asymmetric information about players' types. Assume that the outcome of the game (actions and payoffs) becomes common knowledge after all actions have been taken. Here LoF may worsen the set of Nash equilibria. There may be a set of mutually beneficial transactions that would occur without LoF, but do not occur with LoF because:

1. the proposer does not know for sure whether the other party will accept or reject *and*

2. a high enough probability of rejection can outweigh the expected gains to a successful transaction.

Note that this is in addition to standard problems of asymmetric information: LoF can cause players to reject potential partners even where their *expected* utility from a match is positive. It is also distinct from the "self-image" preservation motive, discussed in Koszegi (2006), that may lead to over- or underconfident task choice.

As we discuss in section 2, there is evidence that a desire not to lose face is a *primal* human concern, perhaps a product of evolutionary factors, or perhaps an automatic internalization of a reputation motive.[1] Thus the LoF may enter into an individual's utility function *directly*. There is a special loss from the combined knowledge that you accepted somebody, but they rejected you. The assumption of a primal LoF leads to a model that bears some resemblance to a *psychological game* (Geanakoplos et al., 1989), in which payoffs depend on beliefs as well as actions. However, in our model, payoffs depend not on beliefs but on "terminal information sets", specifically, the information each player has at the end nodes of the game.[2]

When (e.g.) a woman accepts a man and he rejects her, her material payoffs from this one-shot game are the same no matter what beliefs or information either party has. However, with LoF, her psychic payoffs are lower when *she knows that he knows that* she accepted him and he rejected her. In other words, what the other player knows for sure – the other player's information – is a component of a player's utility function (resembling Battigalli and Dufwenberg (2007)). Unlike in models of fairness (Rabin, 1993; Moreno, 2008) the payoffs *do not* rely on a set of beliefs that are external to the game's structure. Thus, as long as we know the (terminal) information structure, LoF transforms material payoffs into psychological payoffs in a straightforward way. [3] We focus on the primal LoF interpretation: this is particularly relevant to one-shot games where no outside parties observe the results. However, we suspect that many of our results will carry over to a case where the LoF concern can be justified instrumentally. With asymmetric information, as in our model, in many types of dynamic matching and sorting/screening games, her willingness to "accept" another person be taken by others as a negative signal of her type, reducing her utility and/or her continuation value. To reinterpret Groucho Marx "if I am willing to be part of this club, how good can *I* be?" A simple formalization of this in a two-period model is given in the appendix. It is straightforward to derive conditions under which a players' previous "accept" choice hurts her continuation value; however, a complete characterization of equilibria is left for future work.

If loss of face depends on the terminal information sets in this way, i.e., on the information each player has at the end of the game over the game's history, then it can be avoided by changing the information structure so that players *only learn about each other's behavior if they both play Accept.* For example, speed dating agencies often ask men and women to mark the partners who they are interested in, and then inform only those couples who both marked each other. Now, after playing accept, *you* will still be able to infer if you have been rejected, but the *other person* will not know that you accepted them; knowing this, you will not suffer a loss of

---

[1]If reputation concerns are long-term, anticipating the additional short-run pain of losing face may help counteract present-bias as well as overconfidence.

[2]We assume that the structure of this terminal information is common knowledge. We refer to these as "*terminal* information sets" to avoid confusion with the standard terminology, in which information sets are only defined in connection with decision nodes. Our game is less cognitively demanding (on players) than a psychological game, and it is less strategically complex, as our payoffs do not depend on ambiguous higher-order beliefs.

[3]Furthermore, LoF has no obvious interpretation in terms of fairness; since revelation of "who proposed to whom" or "who was kind to whom" occurs after these decisions were made it should have no impact on beliefs about whether a player knew his play was "fair" in the sense of being congruent with the other player's kindness or unkindness.

face. Thus, while your ego-utility can not be preserved, your *face* can be.[4] We call such setups *Conditionally Anonymous Environments* or CAE's.

Our paper proceeds as follows. In section 2 we discuss our concept in more detail and offer intuitive, anecdotal, and academic support for it. We also give a short survey of the related economic literature. In section 3 we describe our baseline setup (similar to a single stage of Chade, 2006), yielding only monotonic equilibria, following the theory of games with strategic complementarities (summarized in Vives, 2005). In section 4 we introduce LoF on one side to the model, and illustrate how this can cause a welfare loss, and how anonymity can improve this outcome. In contrast to Gale and Shapley (1962), we find a second-mover advantage in this context. We present monotone comparative statics as a small amount of LoF is introduced to an environment (an application of Milgrom and Shannon, 1994). We conclude in section 5, considering extensions and discussing policy implications. Longer and more technical proofs are in the appendix, as well as a model illustrating how reputation concerns in a two-stage game may resemble intrinsic LoF.

2 Background

There is abundant psychological evidence that "rejection hurts" (Eisenberger and Lieberman, 2004) and that social ostracism can cause a neurochemical effect that resembles physical pain (Williams, 2007). However, these studies do not distinguish between cases where it is common knowledge that the rejected party has expressed an interest from cases where this is private information. We claim that people fear proposing, and they fear it more when proposals are known. We ask the reader: which of scenarios below would likely cause you more psychological pain? (To keep the example simple, suppose you are romantically interested in women.)

1. A friend or colleague, in whom you have an unexpressed romantic interest, informs you while talking about her tastes that she wouldn't go out with you because you are not "her type." You have no reason to believe that she knows you are interested in her. However, you are fully convinced that she is being honest with you.

2. Without having the conversation in scenario 1, you ask this same person out on a date and she refuses because you are not "her type."

We speculate that the second scenario would be more painful: now both you and she know that you have asked her out and she has refused. Although she may have tried to soften the blow by posing this as a matter of idiosyncratic preference rather than quality, you have lost face, and you are established as her inferior in one sense. In the first case, although you can presume she is not interested in you, and this may hurt your self esteem, she doesn't know you like her, and you have not lost face.

Our speculation here is supported by a variety of previous work. Douglas (1987) "reports eight strategies that individuals reported using to gain affinity-related information from opposite sex others in initial interactions". Bredow et al. (2008) try to capture previous research in presenting the formula $V = f(A \times P)$ for the "strength of the valence of making an overture" to a romantic partner, where A represents attraction and P is the estimated probability that an

---

[4]However, a mechanism designed to match multiple agents on each side (e.g., Gale-Shapley) could be designed to reveal to an individual more or less information about the number of agents who rejected her. An "optimal" matching mechanism might be adjusted to take this into account; this may be an interesting subject for future work.

overture will be accepted. Shanteau and Nagy's (1979) experimental work finds that "when the probability of acceptance is low, people's interest in pursuing a relationship is nil, or nearly nil, regardless of how attracted they are to the person." In the 2005 "Northwestern Speed-Dating Study" on 163 undergraduate students, "participants who desired everyone were perceived as likely to say yes to a large percentage of their speed-dates, and this in turn negatively predicted their desirability" (Eastwick and Finkel, 2008). Baxter and Wilmot (1984) described six types of secret tests used in the delicate dance of "becoming more then friends", e.g., "third-party tests"; these can be seen as motivated by a desire to avoid LoF or loss of reputation (Hitsch et al., 2010).

The fear of LoF is closely related to what psychologists call "rejection sensitivity." For example, London et al (2007) provide evidence from a longitudinal study of middle school students that, for boys, "peer rejection at Time one predicted an increase in anxious and angry expectations of rejection at Time 2." They also find that anxious and angry expectations of rejection are positively correlated to later social anxiety, social withdrawal, and loneliness. In explaining the connection to loneliness, they posit that the rejection sensitive may exhibit "behavioral overreactions" such as "'flight' (social anxiety/withdrawal) or 'fight' (aggression)." It is easy to interpret either of these as a way to choose "reject" in our matching game in order to avoid further loss of face.[5]

Erving Goffman (2005) has written extensively about losing and preserving *face:*

> The term *face* may be defined as the positive social value a person effectively claims for himself by the line others assume he has taken during a particular contact ...The surest way for a person to prevent threats to his face is to avoid contact in which these threats are likely to occur. In all societies one can observe this in the avoidance relationship and in the tendency for certain delicate transactions to be conducted by go-betweens ...

In the context of our paper, Goffman's "avoidance" is essentially preemptive rejection: you cannot be matched with a partner if you don't show up.

In the USA, 17% of heterosexual and 41% of same-sex couples who met in the last ten years met online (Rosenfeld and Thomas, 2012), and the dating industry has been reported to constitute "a $2.1 billion business in the U.S., with online dating services ... representing 53% of the market's value"(MarketData Enterprises, 2012). Internet dating *itself* can be seen as an institution designed to minimize the LoF that comes with face-to-face transactions, allowing people to access a network of potential partners who they are not likely to run into again at the office or on the street. However, going online may not eliminate the LoF; as noted in Hitsch et al. (2010): "If ... the psychological cost of being rejected is high, the man may not send an e-mail, thinking that the woman is 'beyond his reach,' even though he would ideally like to match with her." Perhaps in response to this, several sites and applications on Facebook have introduced the CAE environment, where member A can express interest in member B and member B only finds out about this if B also expresses an interest in A. However, there is a trade-off between preserving face and getting noticed: with thousands of members, each member may only view a fraction of eligible dates, and if A expresses anonymous interest there is no guarantee that B

---

[5]The psychological evidence motivates future work considering heterogeneity in LoF: subjects with observable (e.g., sex, race) and unobservable (e.g., popularity) differences may be more or less concerned with LoF.

will even see A's profile.[6] A Facebook app called "Bang with Friends" promised "Anonymously find friends who are down for the night... we will only show your friends that you're interested if they are too!"; applications and sites such as Tinder have adopted similar methods and garnered millions of regular users.[7]

"Speed dating" events are another relatively recent innovation in the singles scene. These events usually attract an equal number of customers of each gender; men rotate from one woman to another, spending a few minutes in conversation with each. Here there is also an effort to minimize the possibility of LoF. In fact, speed dating agencies often promote themselves on these grounds, e.g., Xpress dating advertises "rejection free dating in a non-pressurized environment"[8]. A recent "Date Matching Methods" patent even claims to "provide[s] methods and means that enable the rapid date matching of single participants, while minimizing the embarrassment of rejection" at an event (Risemberg, 1999). Typically, participants are asked to select who they would like to go on "real dates" with only *after* the event is over. In most cases the agency will only reveal these "proposals" where there is a mutual match, i.e., where both participants have selected each other.[9]

LoF is not limited to the dating world. Both psychological LoF and material losses from publicly observed acceptances and rejections can be seen in many spheres. These concerns are present on both sides of the job market. Within an employment relationship, LoF might affect decisions over how much effort to put in, whether to identify with the firm, and whether to apply for promotions. Akerlof and Kranton (2005) consider a worker's identity as an "outsider" or "insider," and claim that a firm can take steps to change this identity. In their model, "when an employee has [insider] identity, she loses some utility insofar as she deviates from the ideal action of [the firm]," and for an outsider this is reversed. If the differential payoffs derive specifically from her *publicly revealed* actions this could lead to payoffs resembling those of our model. Suppose the firm chooses whether to pay a cost to "approve" of the employee and make her feel like an insider, and the employee simultaneously chooses whether to put in extra effort in the firm's interest. If the firm chooses not to pay this cost, i.e., to "reject" the employee, this will hurt the employee more if she has put in high effort. Thus, employees who fear such an "identity loss" or analogously, feel the LoF from having publicly accepted the firm only to be rejected may only try extra-hard on tasks that they can work on in relative obscurity.

The same authors' (2000) model of social exclusion can be adapted to a LoF interpretation. If being seen "acting white" involves sacrificing Black identity, a Black person may choose not to attempt "admission to the dominant culture" because she is uncertain about the "level of social

---

[6]Online dating has been portrayed as a modern analogue to the traditional "matchmaker," who was able to separately interview prospective mates and their families about their likes and preferences, helping arrange marriages while preserving anonymity. However, the internet and social media cuts both ways. Although the internet affords the opportunity to make connections outside one's usual network, the "gossip network" may grow, increasing reputation concerns.

[7]See <http://www.bangwithfriends.com/>. Accessed 18 April 2013. ¡http://www.latimes.com/business/technology/la-fi-tn-tinder-plus-20141106-story.html¿ accessed 17 January 2015.

[8]<http://www.xpressdating.co.uk/speed_intro.htm>, accessed November 2012.

[9]"Speed dating" institutions have been extended outside the realm of romance and marriage, into forming study groups, "speed networking", "industrial speed dating," and business partnering; these may have been established (in part) to minimize LoF. See: Collins and Goyder (2008), <http://edition.cnn.com/2005/BUSINESS/01/27/speed.networking/index.html>, <http://www.earthwisesociety.bc.ca/TEIP/TEIP/events/08-11-BPS/Speed%20dating%20checklist.pdf>, and <http://www.businesswire.com/portal/site/home/permalink/
?ndmViewId=news_view&newsId=20090622005588&newsLang=en>.

exclusion" she will face; e.g., whether she will be accepted by a school, employer, or White social group. On the other hand, if she can make this attempt anonymously, she can avoid the risk of a public threat to her identity, and also avoid the potential *material* costs of social exclusion. In fact, race-based rejection sensitivity has been found to negatively correlate with measures of African-American students' success at predominantly White universities (Mendoza-Denton et al., 2002). This concern might help justify outreach programs for underrepresented minorities; in effect "asking them first" or letting them know when they will have a high probability of succeeding.

The employer too may be vulnerable to LoF. Cawley (2003), in his guide for economists on the junior job market, writes that he has "heard faculty darkly muttering about job candidates from years ago who led them on for a month before turning them down." This aggravation surely involves LoF in addition to the loss of time. This LoF is recognized by professional recruiters as well: "recruiters lose face when candidates pull out of accepted engagements at the last minute."[10]

For the rejection-sensitive, any economic transaction that involves an "ask" may risk a LoF. This may explain the prevalence of posted prices, aversion to bargaining in certain countries, and the relative absence of neighborhood cooperation, social interaction, consumption and task-sharing in many modern societies (Putnam, 2001). Rejection sensitivity is particularly disabling for sales personnel.[11]Our model may also be important in an archetypal situation where preserving face is valued – the resolution of personal and political disagreements. Neither side may want to make a peaceful overture unilaterally – this can be seen as evidence of admission of guilt or weakness, and may be psychologically painful in itself. Again, where a double-blind mechanism is available, it can resolve this dilemma; if not, our model offers insight into why negotiations often fail. Often, peace talks are made in secret, and only announced if a successful agreement has been reached. This contradicts one of Woodrow Wilson's famous "14 points": "Open covenants openly arrived at", became a principle, according to Eban (1983). However, Eban claims that "the hard truth is that the total denial of privacy even in the early stages ... has made international agreements harder to obtain than ever". Tony Armstrong (1993) analyzed three key cases of international negotiations finding a high degree of secrecy and few participants. "In these secret and private negotiations, assurances and commitments were provided, which were essential for the parties to negotiate 'in good faith'" (Jonsson and Aggestam, 2008).

While economists have previously studied related concepts, to our knowledge none have considered the difference between "mutually-observed acceptance and rejection" and "rejection where only one side knows she was rejected" (and the other side does not know whether or not he was proposed to). Becker (1973) introduced a model of equilibrium matching in his "Theory of Marriage." He considers the surplus generated from marriage through a household production function, and allows the division of output between spouses to be divided ex-ante according to each party's outside option in an efficient "marriage market." Anderson and Smith (2010)

---

[10]Leslie Merrow, Staffing Consultancy Principal, "How to Avoid the Counter-Offer Trap," Friday, August 17, 2007. <http://directsearchalliance.blogspot.com/2007/08/how-to-avoid-counter-offer-trap.html>

[11]"Call reluctance, which strikes both individuals and teams, develops in many forms. Representatives may be 'gun shy' from an onslaught of rejection or actively avoid certain calling situations such as calling high-level decision makers or asking for the order. Call reluctance is the product of fear; fear of failure, fear of losing face, fear of rejection or fear of making a mistake. If the fear perpetuates, productivity suffers." – "Business Services Industry Contests combat call reluctance," Telemarketing & Call Center Solutions, Oct 1996 by Brian J. Geery.

brought reputation into this context, noting "matches yield not only output but also information about types" (but note that *offers* are not observed in their model). Chade (2006) explored a search and matching environment where participants observe "a noisy signal of the true type of any potential mate." He noted "as in the winner's curse in auction theory – information about a partner's type [is] contained in his or her acceptance decision." However, in Chade's model there is only a single interaction between the same man and woman, and outside parties do not observe the results; thus there is no scope for either party's actions to affect their future reputations, nor any direct cost of being rejected.

Our work shares some features of standard models of directed search (e.g., Albrecht et al., 2006), but there are important differences. In the previous employment models, the cost of submitting a job application is unrelated to the probability of success, as well as unrelated to whether the employer knows a candidate has applied. In our model (as applied to this example) the job seeker knows that if there is a high probability they will get the job (even if they do not accept the offer), there is a low probability they will lose face, and thus this cost is low in expectation.

3   Setup and benchmark model

3.1   Agents

The economy is populated by a continuum of individuals on market sides M and F endowed with measure 1 each. An individual $m \in M$ or $f \in F$ is characterized by a binary type $x_j \in \{\ell, h\}$ where $j \in \{m, f\}$; the type is an agent's private information.[12] Let the share of h types be the same on both sides of the market, and denote it by p.[13]

3.2   Matching

Each individual in M is randomly matched to a single individual in F; all matches are chosen by nature with equal probability. Person i obtains a noisy signal s_j about the type x_j of her match j, but does not observe s_i, the signal of her own type that j received. After observing the signals, individuals accept (A) or reject (R) the match. Suppose that acceptance decisions are made sequentially and let males move first.[14] We distinguish two different informational settings: (i) an *Asymmetric Revelation Environment* (ARE) where females observe the action A or R taken by a male but males do not observe females' actions, and (ii) a *Conditionally Anonymous Environment* (CAE) where females do not directly observe the male's action, but only observe whether or not *both parties* have played accept.[15] That is, in the CAE both males and females can infer their match's actions if and only if they themselves play accept.

---

[12] In a previous draft we derived similar results with continuous types under certain functional restrictions. Note that our previous discussion also encompasses one-sided matching; we model a two-sided market (labeled "male" and "female" with apologies for political incorrectness) both because it is relevant to many examples and because it allows us to isolate both direct and indirect effects of a fear of LoF on one side.

[13] This assumption is made for notational simplicity, allowing that the type distributions differ by gender does not affect our results qualitatively.

[14] Modeling a simultaneous move game (available by request) will generate equivalent results, at the expense of considerable notational burden.

[15] We compare the CAE to the *ARE* as the one-way loss of face is easier to model, yields clearer intuition, and is relevant to a sequential game. This is further justified below.

3.3 Payoffs

If both individuals in a matched pair accept they become "married" and each individual's payoff depends positively on the *pizzazz* (see Burdett and Coles, 2006) of their partner: $\tilde{x} \in \{\tilde{\ell}, \tilde{h}\}$. ℓ types have pizzazz $\tilde{\ell}$ and h types have pizzazz $\tilde{h}$, where $0 < \tilde{\ell} < \tilde{h}$. Types (and thus pizzazz) become fully observable during the marriage. Marriage payoffs for a match (m, f) are given by $u_m(\tilde{x}_f) = \tilde{x}_f$ and $u_f(\tilde{x}_m) = \tilde{x}_m$. That is, payoffs are linear in the match's type, which implies that total surplus does not depend on the precise assignment of types, but only on the number of marriages formed.[16] Agents who remain solitary obtain a payoff of $u_j(\tilde{x}_j) = \delta \tilde{x}_j$ for $j \in \{m, f\}$ with $\delta < 1$. For intuition, suppose types represent productivity, production is shared by the married couple, and those who are more productive alone are also more productive in a marriage.[17] Therefore ℓ types always prefer a marriage to remaining alone. To make the model non-trivial, we suppose that h's prefer to remain unmarried to marrying an ℓ, i.e., high types have a good outside option:

Assumption 1. $\delta \tilde{h} > \tilde{\ell}$.

The crucial assumption here is that there are only two types; allowing for a larger, finite type space would be straightforward. In summary, homogenous marriages benefit both partners and mixed marriages benefit ℓ's more than they hurt h's.[18] The payoff matrix of the game played by a randomly matched pair (m, f) is thus

$$
\begin{array}{c|cc}
{}_f\!\diagdown^m & A & R \\
\hline
A & \tilde{x}_f, \tilde{x}_m & \delta \tilde{x}_m, \delta \tilde{x}_f \\
R & \delta \tilde{x}_m, \delta \tilde{x}_f & \delta \tilde{x}_m, \delta \tilde{x}_f
\end{array}
\tag{1}
$$

While our setting does not allow for a generic analysis of matching games, it captures a large set of interactions where loss of face may be relevant. Our assumptions embody agreed-upon preferences over a partner's type – partners are better or worse along a single dimension, although this may be a reduction of several characteristics. This is plausible, at least to first order, for measures such as desirability in the marriage market and productivity in a market for business partners.

3.4 Signals

Individuals in a matched pair each obtain a signal $s \in [\underline{s}, \overline{s}]$ of the other agent's type. Signals are drawn independently and their distribution depends on the type of the sender: type x's $(x \in \{\ell, h\})$ signal is distributed according to $F_x(s)$ with continuously differentiable density $f_x(s)$. Suppose that the signal is informative in the sense that $f_\ell(s)$ and $f_h(s)$ satisfy the monotone likelihood ratio property, i.e.,

Assumption 2. $f_h(s)/f_\ell(s) > f_h(s')/f_\ell(s')$ *for all $s > s'$ where defined.*

---

[16]We see this assumption as a reasonable benchmark, representing the knife-edge between increasing and decreasing differences. Allowing super- or sub-modularities in the marriage production function may be interesting for future research.

[17]This assumption might also be justified through interpreting the payoff to no match as the continuation value in an indefinitely repeated version of the matching game, as in, e.g., Adachi (2003). In Chade (2006) this value increases in type as higher-type players tend to have higher signals and other players accept them more often.

[18]This follows from the larger total payoff, as $\tilde{h} + \tilde{\ell} > \delta(\tilde{h} + \tilde{\ell})$.

We assume that signals have limit distributions, i.e., that observing the best (worst) signal implies that the type is h (ℓ), i.e.,

Assumption 3. $f_h(\underline{s}) = 0$, $f_\ell(\underline{s}) > 0$, $f_\ell(\overline{s}) = 0$, and $f_h(\overline{s}) > 0$.

Assuming that the probability of a high (low) type converges to one (zero) will ensure that game has an interior equilibrium (signal thresholds for accepting a match will be interior), and it is crucial for lemma 1, as further discussed below.[19]

We will look for Perfect Bayesian Equilibria and consider tatonnement stability, i.e., stability with respect to the iterative responses to deviations or "cobweb dynamics" (see Hahn, 1962, Dixit, 1986, and Vives, 2005).[20]

3.5 Benchmark model: no loss of face

To illustrate the setup we now solve for the equilibrium behavior in the matching game without loss of face, which is equivalent to the CAE. Since $\delta < 1$, low types will always prefer any marriage to remaining alone. Therefore low types find it always (at least weakly) optimal to accept.

There is always a "trivial coordination failure" equilibrium where both players always reject, since if i's match rejects with certainty, then either accepting or rejecting yield payoff $\delta \tilde{x}_i$. However, this equilibrium is neither tatonnement-stable nor trembling-hand perfect; if low (or high) types reduce their cutpoint slightly or tremble,[21] thus playing accept with a positive probability, other low types will strictly prefer to play accept, i.e., set their cutpoint at $\underline{s}$. Thus, in a stable equilibrium low types must always accept with some minimum probability $\varepsilon > 0$, which in turn, given assumption 3, implies that high types must at least sometimes reject. This implies that the acceptance behavior of players of type h and ℓ differs in equilibrium, implying that the event of being accepted also conveys some information about the match's type (the "acceptance curse" in Chade, 2006). Let $q_j(x; x')$ for $j \in \{m, f\}$ denote the resulting equilibrium probability that a type x on side j accepts a match, given the match is of type x'.[22] Since ℓ types always find it profitable to accept, $q_j(\ell, x') = 1$ for $x' \in \{\ell, h\}$ and $j \in \{m, f\}$.

Note that the above "trivial coordination failure" is distinct from the case where low types always accept and high types always reject, which we call the Coordination Failure Equilibrium or "CFE"; we return to this below.

For an individual i of type h in a match (i, j), playing R yields a payoff $\delta \tilde{h}$, whereas playing A either yields x_j, if j accepted, or $\delta \tilde{h}$ if j rejected. That is, the decision to accept or reject only affects the payoff in the case that the match accepted. High types of both genders (whether or not they observed their match's action) find it (weakly) profitable to accept after observing a

---

[19]Alternatively, we could consider the alternate assumption of overlapping supports i.e., for all $s \in [\underline{s}, \overline{s}]$ $f_h(s) > 0$ if and only if $f_\ell(s) > 0$. This alternative assumption would imply that we could not rule out equilibria where high types do not respond to the signal and instead "always accept" or "always reject," and these could be stable. However, even under overlapping supports, if we were to focus on "responsive" equilibria where high types have interior thresholds, our remaining results carry over. Details are available by request.

[20]We will also mention when our results hold under the trembling-hand perfection refinement.

[21]As individuals do not observe the signals of their own type, a reduced cutpoint on the other side will be perceived in the same way and elicit the same response as a random tremble.

[22]Note that this expression incorporates the acceptance curse because this allows for high types to accept less, i.e., $q(h; h) < 1$, and being accepted means a higher likelihood that one's match is a low type.

signal s if and only if the expected payoff from accepting meets or exceeds the outside option:

$$\frac{pf_h(s)}{(1-p)f_\ell(s)+pf_h(s)}[q_j(h,h)\tilde{h}+(1-q_j(h,h))\delta\tilde{h}]+\frac{(1-p)f_\ell(s)}{(1-p)f_\ell(s)+pf_h(s)}\tilde{\ell}\geq\delta\tilde{h}, \quad (2)$$

where $j=m,f$. If we subtract $\delta\tilde{h}$ from both sides of the above inequality and rearrange terms, we see that an h accepts if the gain from marrying another h outweighs the loss from marrying an ℓ in expectation conditional on the signal and considering equilibrium play, i.e., if:[23]

$$pf_h(s)[q_j(h,h)\tilde{h}+(1-q_j(h,h))\delta\tilde{h}-\delta\tilde{h}]\geq(1-p)f_\ell(s)(\delta\tilde{h}-\tilde{\ell}). \quad (3)$$

Note that the overall probability of being accepted as a function of types, $q_j(x';x)$ does not depend on the signal s that the player observes, as signals are drawn independently and individuals do not observe the signals of their own type. Because of the monotone likelihood property, (2) implies that an agent i of type h accepts if and only if the signal s that i observes is sufficiently high. Hence, there is $\hat{s}\in[\underline{s},\overline{s}]$ such that an agent of type h finds it profitable to accept if and only if he observes $s\geq\hat{s}$. That is, both males and females use threshold strategies, accepting only if the signal exceeds their threshold, $\hat{s}_m$ and $\hat{s}_f$ respectively. This implies that an agent $j\in\{m,f\}$ of type h accepts an agent of type x with probability

$$q_j(h,x)=1-F_x(\hat{s}_j), \text{ with } j\in\{m,f\}.$$

Using this on (2) implies that, given a cutoff value $\hat{s}_j$ on the other market side, a high type's optimal reply is to accept if

$$\frac{f_h(s)}{f_\ell(s)}\geq\frac{1-p}{p}\frac{\delta\tilde{h}-\tilde{\ell}}{(1-\delta)\tilde{h}}\frac{1}{1-F_h(\hat{s}_j)}, \quad (4)$$

and to reject otherwise. We can thus define the high types' best reply in any stable equilibrium: accept if and only if $s\geq\hat{s}_i^*(\hat{s}_j)$, where $i\neq j\in\{m;f\}$ and

$$\hat{s}_i^*(\hat{s}_j): \frac{f_h(\hat{s}_i^*)}{f_\ell(\hat{s}_i^*)}=\frac{1-p}{p}\frac{\delta\tilde{h}-\tilde{\ell}}{(1-\delta)\tilde{h}}\frac{1}{1-F_h(\hat{s}_j)}. \quad (5)$$

Continuity of the signal and its distribution and the monotone likelihood property ensure that the best response $\hat{s}_i^*(.)$ exists and that $\hat{s}_i^*(.)$ weakly increases in its argument. The latter implies that thresholds are strategic complements. Combined with the symmetry of the game in types this implies that $\hat{s}_m=\hat{s}_f:=\hat{s}^*$ in a Nash equilibrium (supposing the contrary leads quickly to a contradiction). As noted above, $\hat{s}^*>\underline{s}$ in any stable equilibrium. Since agents' actions do not affect other agents' information sets, beliefs are always formed according to Bayes' rule, and the issue of out-of-equilibrium beliefs will not arise.

Totally differentiating and rearranging terms (see footnote) we see the slope of a high-type

---

[23]Rearranging, $\frac{pf_h(s)}{(1-p)f_\ell(s)+pf_h(s)}[q_j(h,h)\tilde{h}+(1-q_j(h,h))\delta\tilde{h}-\delta\tilde{h}]+\frac{(1-p)f_\ell(s)}{(1-p)f_\ell(s)+pf_h(s)}(\tilde{\ell}-\delta\tilde{h})\geq 0$ $\Leftrightarrow pf_h(s)[q_j(h,h)\tilde{h}+(1-q_j(h,h))\delta\tilde{h}-\delta\tilde{h}]+(1-p)f_\ell(s)(\tilde{\ell}-\delta\tilde{h})\geq 0$, $\Leftrightarrow$ inequality 3.

player i's best response function must follow:[24]

$$\frac{\partial \hat{s}_i}{\partial \hat{s}_j} = \frac{f_h(\hat{s}_i)f_h(\hat{s}_j)}{\frac{\partial f_h(\hat{s}_i)}{\partial s}(1 - F_h(\hat{s}_j)) - \frac{1-p}{p}\frac{\delta\tilde{h}-\tilde{\ell}}{(1-\delta)\tilde{h}}\frac{\partial f_\ell(\hat{s}_i)}{\partial s}}. \tag{6}$$

A sufficient condition for an interior stable equilibrium to exist is

$$(f_h(\overline{s}))^2 > -\frac{\partial f_\ell(\overline{s})}{\partial s}\frac{1-p}{p}\frac{\delta\tilde{h}-\tilde{\ell}}{(1-\delta)\tilde{h}}. \tag{7}$$

Condition 7 ensures that that a high type's reaction function has a slope greater than unity at $\overline{s}$ (which also implies that the CFE is not stable). Condition 7 would be implied by $\frac{\partial f_\ell(\overline{s})}{\partial s} = 0$, i.e., if the low type's signal distribution becomes flat at $\overline{s}$.[25] Condition 7 will also hold if the underlying share of high types is sufficiently great; e.g., for a linear density on $[0, 1]$ condition 7 holds if $p > \frac{\delta\tilde{h}-\tilde{\ell}}{2\tilde{h}(1-\delta)+\delta\tilde{h}-\tilde{\ell}}$.

Note that the slope of the best response function (6) must be zero at $\hat{s}_j = \underline{s}$ because of the continuous limit distributions ($f_h(\underline{s}) = 0$). Further recalling $\hat{s} > \underline{s}$, we see that if condition 7 holds, then the best response function must intersect the 45 degree line at least once, so there will an interior equilibrium, as shown in Figure 1. Since the best response function is increasing, at least one intersection point will have a slope below one and thus correspond to a stable equilibrium. For a unique equilibrium we would further require the best response function to be convex. If condition 7 does not hold then the CFE will be stable, and we may not have an interior equilibrium.
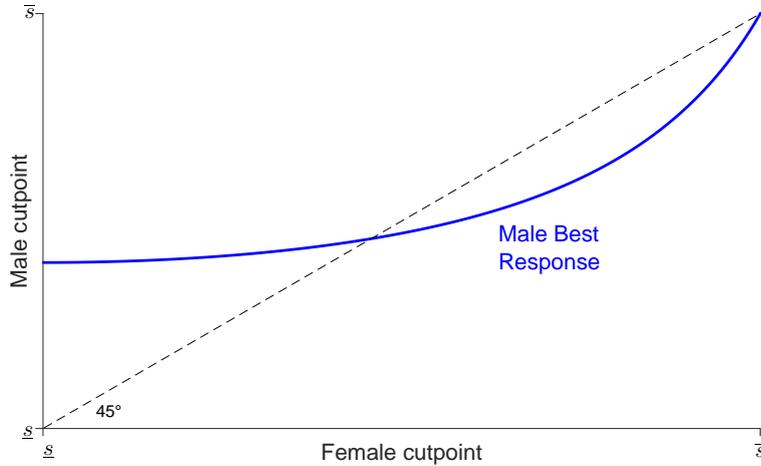

Figure 1: Male high type's best response to high type female cutpoints; condition 7 holds here

We summarize the above results for the game without loss of face in the following lemma.

Lemma 1. *If condition 7 holds then at least one interior, stable equilibrium exists, and in any stable equilibrium*

---

[24]Rearranging equation 5, i's optimal cutpoint $\hat{s}_i^*(\hat{s}_j)$ is defined by $f_h(\hat{s}_i^*)(1 - F_h(\hat{s}_j)) = \frac{1-p}{p}\frac{\delta\tilde{h}-\tilde{\ell}}{(1-\delta)\tilde{h}}f_\ell(\hat{s}_i^*)$. Totally differentiating yields: $\frac{\partial f_h(\hat{s}_i^*)}{\partial s}(1 - F_h(\hat{s}_j))\Delta\hat{s}_i^* - f_h(\hat{s}_i^*)f_h(\hat{s}_j)\Delta\hat{s}_j = \frac{1-p}{p}\frac{\delta\tilde{h}-\tilde{\ell}}{(1-\delta)\tilde{h}}\frac{\partial f_\ell(\hat{s}_i^*)}{\partial s}\Delta\hat{s}_i^*$. Rearranging yields: $\left(\frac{\partial f_h(\hat{s}_i^*)}{\partial s}(1 - F_h(\hat{s}_j)) - \frac{1-p}{p}\frac{\delta\tilde{h}-\tilde{\ell}}{(1-\delta)\tilde{h}}\frac{\partial f_\ell(\hat{s}_i^*)}{\partial s}\right)\Delta\hat{s}_i^* = f_h(\hat{s}_i^*)f_h(\hat{s}_j)\Delta\hat{s}_j$. The usual argument then yields differentiability of the best response function and the slope of the reaction function in equation 6.

[25]This would also ensure that the CFE is not trembling-hand perfect, since any tremble means that a high type's dominant strategy will be to accept any signal that cannot come from a low type.

1. low types always accept,

2. high types use symmetric cutoff strategies, accepting if $s > \hat{s}_m = \hat{s}_f := \hat{s}^*$, and

3. $\hat{s}^* \in (\underline{s}, \overline{s})$.

Simple calculations yield a corollary to this result:

Corollary. *Where condition 7 holds, expected payoffs for types ℓ and h in a stable equilibrium of the game without loss of face (or for any strategy profile where low types always accept) are*

$$v(\ell) = \delta\tilde{\ell} + p(1 - F_\ell(\hat{s}^*))(\tilde{h} - \delta\tilde{\ell}) + (1 - p)(\tilde{\ell} - \delta\tilde{\ell})$$
$$v(h) = \delta\tilde{h} + p(1 - F_h(\hat{s}^*))^2(\tilde{h} - \delta\tilde{h}) - (1 - p)(1 - F_\ell(\hat{s}))(\delta\tilde{h} - \tilde{\ell}). \tag{8}$$

Note $v(h) > v(\ell)$. The number of marriages is $(1-p)^2 + 2p(1-p)(1 - F_\ell(\hat{s}^*)) + p^2(1 - F_h(\hat{s}^*))^2$, which strictly decreases in $\hat{s}^*$.

Intuitively, an ℓ will not marry (and will thus get $\delta\tilde{\ell}$) unless he meets another ℓ or fools an h. An h will marry only if she meets another h and they both send very positive signals, or if she is fooled by an ℓ (i.e., she meets an ℓ who sends a high enough signal).

In the above game without loss of face the order of play does not affect outcomes. Since both players' actions impact payoffs only in the case that the other player plays accept, their best responses do not change if they learn whether or not the other player *did* in fact accept; the same acceptance curse occurs in either case.

4 Intrinsic Loss of Face in an Asymmetric Revelation Environment

Loss of face as described in section 2 is an intrinsic psychological pain, which can only matter if a player's potentially embarrassing action is observed by the other player. Therefore we define loss of face as follows.

Definition 1. *A player j who suffers from intrinsic loss of face suffers a loss L when*

1. j played accept. j knows that his match, player k, played reject, and

2. j knows that k knows that j played accept.[26]

Since LoF results from the *common knowledge* (or at least the higher order beliefs described above) of one party accepting and the other rejecting, to model LoF we need to make payoffs depend not only on actions, but also on the information players hold at the end of the game. These *terminal information sets* for players m and f are defined as standard information sets, but they are not at a decision node: they characterize a player's knowledge about the complete history of the game after all actions have been taken.[27]

In the game defined above terminal information sets will depend on the information environment in place (this is graphically depicted in figure 3 in the appendix). Denote an action tuple

---

[26]Point 2 is necessary for a *primal* LoF but not for the reputational LoF we model in Appendix B.1; my reputation and future payoffs may suffer whether or not I know that my decision is observed.

[27]For the primal LoF, this is related to the Value of Useless Information model introduced by Alaoui (2010).

by $(a_m, a_f) \in \{A; R\}^2$. If both players in a match only observe their own actions and whether or not there is a marriage (the conditionally anonymous environment, CAE), both players' information sets are (A, A), $\{(R, A), (A, R)\}$, or (A, R). This implies that females cannot distinguish between action profiles (A, R) and (R, R), and males cannot distinguish between (R, A) and (R, R). Therefore there is no loss of face under the CAE, and equilibrium behavior under the CAE is given as in the previous section (see especially lemma 1 and its corollary).

As noted earlier, having only one side vulnerable to LoF will allow us to separately consider both direct and indirect effects, and will be relevant to the comparison between sequential decision environments and the CAE. Here we will assume that the males move first; i.e., consider "males" as synonymous with "first-movers" or "the side vulnerable to LoF."

If females observe males' actions in a match, but not vice versa (the asymmetric revelation environment, ARE), possible terminal information sets are (R, A), (R, R), (A, A), or (A, R) for females, and (A, A), (A, R), and $\{(R, A); (R, R)\}$ for males.[28] That is, it is common knowledge that females observe the action profile (A, R), whereas males cannot distinguish (R, A) and (R, R). Therefore there is loss of face for males when the action profile (A, R) is played, but there is no loss of face for females under the ARE.[29]

Under the ARE the payoff matrix (1) changes into:

$$
\begin{array}{c|cc}
{}_m\;\diagdown^{\;f} & A & R \\
\hline
A & \tilde{x}_f, \tilde{x}_m & \delta\tilde{x}_f - L, \delta\tilde{x}_f \\
R & \delta\tilde{x}_f,\, \delta\tilde{x}_m & \delta\tilde{x}_f,\, \delta\tilde{x}_m
\end{array}
\quad . \tag{9}
$$

As above individuals' acceptance decisions will depend on the inference they make about their match's type given the signal and given the event of being accepted. Because only males may incur loss of face, the decision problems of males and females now differ.

4.1 Female behavior in the ARE

We begin with the females' decision. Since females do not suffer from loss of face the analysis from section 3.5 carries over: $\delta > 0$ implies again that low-type females find it always profitable to accept. As before, denote by $q_m(h, h)$ and $q_m(\ell, h)$ the probabilities that male h and ℓ types, respectively, accept female h types in equilibrium (and q_f is defined similarly for females accepting males). Female h types find accepting profitable if, given the observed signal s, the expected payoff from accepting (conditional on being accepted) exceeds the outside option. Thus, restating inequality 2 for females,

$$
\frac{p f_h(s)}{(1-p) f_\ell(s) + p f_h(s)} [q_m(h, h)\tilde{h} + (1 - q_m(h, h))\delta\tilde{h}] + \frac{(1-p) f_\ell(s)}{(1-p) f_\ell(s) + p f_h(s)}\tilde{\ell} \geq \delta\tilde{h}. \tag{10}
$$

Since $q_m(h, h)$ and $q_m(\ell, h)$ do not depend on the signal that the female player observes, the monotone likelihood property implies, as in the benchmark model, that a female of type h accepts if and only if she observes the signal $s \geq \hat{s}$ for some threshold $\hat{s} \in [\underline{s}, \overline{s}]$.

---

[28] Examining the remaining two possible informational environments is straightforward and does not yield any additional insights.

[29] Equivalently, even if both females and males had four terminal information sets, if the males choose first and the females observe this and choose second, a female would never accept unless the male did, and thus the outcomes would be *as if* she were not vulnerable to LoF.

4.2 Male behavior in the ARE

Given female behavior, males of type x who observe a signal s find it profitable to accept if and only if the expected payoff from accepting (conditional on being accepted), minus the expected loss of face when rejected, exceeds the outside option:

$$\underbrace{\frac{(1-p)f_\ell(s)}{(1-p)f_\ell(s)+pf_h(s)}}_{Pr(x_f=\ell|s)}\tilde{\ell} + \underbrace{\frac{pf_h(s)}{(1-p)f_\ell(s)+pf_h(s)}}_{Pr(x_f=h|s)} \times \left[\underbrace{(1-F_x(\hat{s}_f))\tilde{h}}_{Pr(h\text{ fem. accepts})\times\text{marr. util.}} + \underbrace{F_x(\hat{s}_f)(\delta\tilde{x}-L)}_{Pr(h\text{ fem. rejects})\times\text{reject. util.}} \right] \geq \delta\tilde{x}. \tag{11}$$

Collecting terms, a male of type x accepts if

$$(1-p)f_\ell(s)(\tilde{\ell}-\delta\tilde{x}) + pf_h(s)\left[F_x(\hat{s}_f)(\delta\tilde{x}-L-\tilde{h})+\tilde{h}-\delta\tilde{x}\right] \geq 0.$$

Stating this condition in terms of likelihood ratios we have

$$\frac{1-p}{p}\frac{f_\ell(s)}{f_h(s)}(\tilde{\ell}-\delta\tilde{x}) - F_x(\hat{s}_f)(L+\tilde{h}-\delta\tilde{x}) + \tilde{h}-\delta\tilde{x} \geq 0. \tag{12}$$

Now the monotone likelihood ratio property implies that the left-hand side strictly *decreases* in s for low type males and strictly *increases* in s for high males, implying any cutoff thresholds must be floors for high-type females and *ceilings* for low-type males. Furthermore, male low types accept with probability one if they would prefer to accept when matched with a high-type female (i.e., if the expected value of the marriage gain exceeds the expected loss of face from possible rejection), i.e.,

$$(1-F_\ell(\hat{s}_f))(\tilde{h}-\delta\tilde{\ell}) \geq F_\ell(\hat{s}_f)L, \tag{13}$$

while male high types *reject* with probability one if they would prefer to reject even when matched with a high-type female, i.e.,

$$(1-F_h(\hat{s}_f))(\tilde{h}-\delta\tilde{h}) \leq F_h(\hat{s}_f)L. \tag{14}$$

This implies the following lemma, derived in the appendix.

Lemma 2. *With loss of face in an ARE males use threshold strategies; ℓ types use ceilings, h types use floors.*

1. *Male h types play "accept iff $s_f \geq \hat{s}_h$" with $\hat{s}_h = \overline{s}$ if (14) holds; otherwise $\hat{s}_h$ is interior and is defined by*

$$\frac{f_h(\hat{s}_h)}{f_\ell(\hat{s}_h)}\frac{p}{1-p}\frac{(1-\delta)\tilde{h}}{\delta\tilde{h}-\tilde{\ell}} = \frac{1}{1-F_h(\hat{s}_f)(1+L/((1-\delta)\tilde{h}))}. \tag{15}$$

2. *Male ℓ types play "accept iff $s_f \leq \hat{s}_\ell$" with $\hat{s}_\ell = \overline{s}$ if (13) holds; otherwise $\hat{s}_\ell$ is interior and is defined by*

$$\frac{f_h(\hat{s}_\ell)}{f_\ell(\hat{s}_\ell)}\frac{p}{1-p}\frac{\tilde{h}-\delta\tilde{\ell}}{\delta\tilde{\ell}-\tilde{\ell}} = \frac{1}{1-F_h(\hat{s}_f)(1+L/(\tilde{h}-\delta\tilde{\ell}))}. \tag{16}$$

3. *Female types' best responses are as in the benchmark case.*

As in the benchmark case, male high types attempt to avoid marrying female low types and therefore screen for female high types using the signal. However, unlike the benchmark case, with sufficiently severe loss of face, male low types attempt to avoid being rejected and use the signal to screen for female *low* types. Thus loss of face can make low types behave as "reverse snobs", rejecting only those matches who send higher signals.
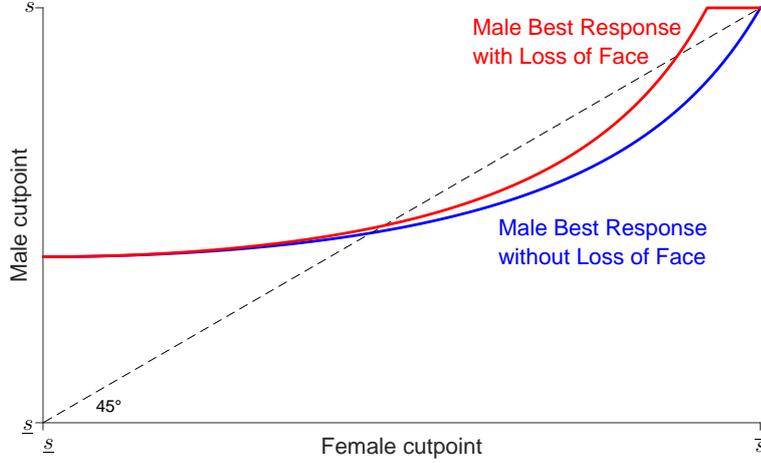

Figure 2: Male high-type's best response to (high type) female cutpoints, with and without loss of face; condition 7 holds here

4.3 Properties of an equilibrium in the ARE

We focus here on the case where an interior equilibrium exists with or without LoF (implying that the CFE is unstable without LoF). Since males continue to use threshold strategies we can now pin down the high-type females' threshold $\hat{s}_f$ (recalling that female low types always accept). Adjusting to allow low type males to set a ceiling $\hat{s}_\ell$, after observing a signal s a female high type accepts if and only if

$$p(1 - F_h(\hat{s}_h))f_h(s)(\tilde{h} - \delta\tilde{h}) \geq (1 - p)F_h(\hat{s}_\ell)f_\ell(s)(\delta\tilde{h} - \tilde{\ell}).$$

This implies that where

$$(1 - F_h(\hat{s}_h))(\tilde{h} - \delta\tilde{h}) < 0, \tag{17}$$

a female high type will never accept (i.e., $\hat{s}_f = \overline{s}$), which can only hold if male high types never accept, i.e., where $\hat{s}_h = \overline{s}$; i.e., this is the CFE.

Otherwise her cutoff $\hat{s}_f$ will be interior and given by

$$\frac{f_h(\hat{s}_f)}{f_\ell(\hat{s}_f)} \frac{p}{1 - p} \frac{(1 - \delta)\tilde{h}}{\delta\tilde{h} - \tilde{\ell}} = \frac{F_h(\hat{s}_\ell)}{1 - F_h(\hat{s}_h)}. \tag{18}$$

Comparative statics

We next consider how the equilibrium changes around the benchmark $L = 0$, as we add a small loss of face term, $L > 0$. For $L = 0$, (the benchmark case) low types always accept, and an interior threshold ($\hat{s}^*$) for female high types satisfies (5), that is,

$$\frac{f_h(\hat{s}^*)}{f_\ell(\hat{s}^*)} \frac{p}{1 - p} \frac{(1 - \delta)\tilde{h}}{\delta\tilde{h} - \tilde{\ell}} = \frac{1}{1 - F_h(\hat{s}^*)}. \tag{19}$$

Perturbing this with a small $L > 0$, we examine the behavior of the system of equations (15, 16, and 18) in the neighborhood of $\hat{s}^*$. For a sufficiently small $L > 0$ condition 13 holds, thus all low types accept, but the high types' optimal thresholds are perturbed. Noting this yields the following statement (proof in the appendix).

Proposition 1. *Suppose that $L = 0$ and that there is an interior equilibrium. Here $\hat{s}^*$ will be implicitly defined by (19). Then, for a small increase in L of $\epsilon > 0$ low males still always accept, i.e. $\hat{s}_\ell = \overline{s}$, and the equilibrium cutoffs for high females and high males move in the same direction. For any interior equilibrium the cutoffs will increase if the local slopes of the reaction functions are below 1, i.e., in a stable equilibrium; they will decrease if these slopes exceed 1, i.e., in an unstable equilibrium.*

For intuition, consider that for equilibrium dynamics, becoming more selective by increasing one's cutoff has a twofold effect on the expected quality of a marriage partner. First, there is a screening effect, increasing the expected quality of a match *holding constant* the acceptance behavior of the other gender. Second there is a supply effect in the opposite direction: if one side becomes more selective, then the other side will react by also becoming more selective, implying a greater acceptance curse on both sides.[30]

The screening effect response to LoF is straightforward. Holding the females' cutoffs constant, the risk of LoF makes playing accept less attractive for males. Thus a high type male would require a greater expected payoff from marriage to get him to play accept, and thus he might raise his signal. This would make the female raise her signal as well (although as noted, in a stable equilibrium, by not as much).

However, if we restrict our consideration to stable equilibria, the counter-intuitive response can be ruled out. As noted in the above proposition, starting from any stable equilibrium LoF makes all high types *more* selective, i.e., $\hat{s}_h > \hat{s}_f > \hat{s}^*$.

Corollary. *(to Proposition 1) Suppose that $L = 0$. For any stable interior equilibrium, for a small increase in L of $\epsilon > 0$ low males still always accept, i.e. their ceiling is $\hat{s}_\ell = \overline{s}$, and the equilibrium cutoffs for high females and high males increase; hence there are fewer marriages and there is lower net welfare.*

Which side is affected more?

Consider: Is the side that bears the loss of face, here the male side, more affected than the other side? Note first that male ℓ types are always at least as selective as female ℓ types, as the latter always accept. For high types, the possibility of losing face may make males more reluctant to accept than females. On the other hand, this effect will increase the females' acceptance curse: it will decrease the probability that, given a female is accepted, her match was a high-type male; thus this will make high-type females more cautious. The following lemma establishes that the first effect dominates.[31]

---

[30]The equilibrium tradeoff between screening and the acceptance curse was present without LoF. However, in the ARE, LoF makes playing accept less attractive for males, and this effect is stronger the more that females reject, hence a steeper reaction function.

[31]Proposition 1 already found that, in a stable interior equilibrium, the market side that incurs a *small* loss of face becomes relatively more selective, $\hat{s}_m > \hat{s}_f > \hat{s}^*$. This lemma ensures that this holds at least weakly in *any* stable equilibrium.

Lemma 3. *In any equilibrium in an ARE with Loss of Face male h types are more selective than female h types, i.e., $\hat{s}_f \leq \hat{s}_h$, and strictly so if the equilibrium is not a coordination failure, i.e., if $\hat{s}_h < \bar{s}$.*

Proof. Let $L > 0$. Suppose that $\hat{s}_f \geq \hat{s}_h$. Then the monotone likelihood property and equations (18, 15 and 16) imply that

$$F_h(\hat{s}_\ell) \geq \frac{1 - F_h(\hat{s}_h)}{1 - F_h(\hat{s}_m)(1 + L/((1-\delta)\tilde{h}))} > 1,$$

a contradiction. $\qquad\square$

In other words, considering high types of both genders, unless LoF induces a coordination failure, the gender that is vulnerable to LoF will be more "snobbish" than the gender that is sheltered from it.

This has an interesting extension. Note that when L is small enough such that $\hat{s}_\ell = \bar{s}$, the probability that a high male marries "below his station": $1 - F_\ell(\hat{s}_h^*)$, is less than $1 - F_\ell(\hat{s}_m^*)$, the probability that a high female does so. Thus, in any stable equilibrium in an ARE with a small amount of LoF: (i) high males marry less often than high females but get better spouses on average, and (ii) low males marry more often than low females and get better spouses on average; thus for low types, a small amount of LoF on one side hurts the *other* side more.

Note that only the market side that *proposes* may incur loss of face, suggesting a contrast from Gale and Shapley (1962).[32]

4.4   Coordination failure

Again there is the possibility of an equilibrium where all h types reject with certainty. However, unlike in section 3.5, with male loss of face $L > 0$, under the ARE this equilibrium is *always* stable. The reason is that with loss of face male h types *strictly* prefer to reject with certainty whenever female h types reject with sufficiently high probability. The intuition is straightforward: in the CFE, high males will not respond to a tremble (nor a reduction in the cutpoint) where a small ϵ share of females play accept. Even if they are almost-surely facing a high type female, the $(1-\epsilon)$ chance of losing face outweighs the ϵ chance of marrying high.

Lemma 4. *Again, there is a CFE where h types on both sides play "reject with probability 1." When one (or more) side is vulnerable to LoF this is stable (and trembling hand perfect). In this CFE male ℓ types accept if they observe $s \leq \hat{s}_\ell$, where $\bar{s} > \hat{s}_\ell > \underline{s}$.*

Indeed there is a possibility of coordination failure arising as one moves from the benchmark setting, where loss of face will never occur, to the ARE with a positive loss of face term.[33] While for the CAE or for $L = 0$ an interior equilibrium is reached whenever condition 7 holds (the coordination failure equilibrium is not stable), the coordination failure is always stable for positive L. In fact, the coordination failure is the unique equilibrium for a large enough increase in L; this follows immediately from condition (14). The following corollary compares

---

[32]In Gale and Shapley (1962) the *proposers* in their deferred-acceptance algorithm are able to secure better matches in equilibrium. For instance, if the males propose to the women the men-optimal matching outcome will attain. This method is often used in practice, for instance the student-optimal algorithm in school choice. Our setup suggests a potential behavioural cost to students if LoF is relevant in this domain, unless a CAE is used.

[33]This would also hold for a Full Revelation Environment (FRE), where both sides are vulnerable to LoF; details by request.

the coordination failure to the benchmark outcome without loss of face; details are in the appendix.

Corollary. *(to lemma 4) Compared to an interior equilibrium allocation of the benchmark model without loss of face, the coordination failure equilibrium in the ARE with loss of face induces*

1. *a lower overall marriage rate,*

2. *lower aggregate surplus (even without directly including loss of face in the surplus calculation),*

3. *and lower expected surplus (again, even without subtracting the LoF) for both types of both genders.*

4.5 Summation

This section has demonstrated that the presence of loss of face can worsen outcomes. Where condition 7 holds, our comparative statics results (Proposition 1) demonstrate that LoF leads to fewer marriages and thus lower payoffs in an interior equilibrium. Here, LoF also makes the CFE stable, and thus may lead to an even worse outcome, where high types never marry. If the aforementioned condition fails to hold then the CFE may be the only stable equilibrium with or without LoF; LoF still would impose a cost (on low types), but only a direct one, with no impact on marriage outcomes.

This has a practical implication for the real-world: even if LoF is present, the equilibrium in the benchmark setup can be reached by adequate design of the matching game. Consider a case of severe loss of face, that is, $L > 0$ sufficiently great, such that coordination failure arises as the only equilibrium in an ARE, and low males set a cutoff ceiling.[34] Changing the setup to a CAE, where females only observe whether a marriage occurs, but not the male's action, A or R, means that now high types set an interior cutoff, and low types always accept in any stable equilibrium. Note that if only males are vulnerable to LoF and females are not ($L_m > 0$ but $L_f = 0$), it would be sufficient to reverse the order and let the females choose first. This may be helpful as the CAE will typically require a trustworthy intermediary, man or machine, while reversing the order would not.

5 Conclusions and suggestions for future work

Our work has potential normative implications. Some mechanisms and policies may be more efficient than others in the presence of LoF concerns, and firms and policymakers should take this into account. Although setting up conditionally anonymous environments may take some administrative effort, and may require a third-party monitor, we suspect that there are many cases in which it will lead to more and better matches and improve outcomes. Consider, for example, the matching of advisors and students in a Ph.D. program. A "tick box system" might work, but we imagine that some might be reluctant to participate in such an impersonal and rigid system. More generally, the use of a knowledgable, reliable, and discrete intermediary, might be more effective. Our paper motivates the use of such "matchmakers" in many contexts.[35]

---

[34]Note that for L extremely large, the male ℓ types' ceiling approaches $\underline{s}$, and the rate of marriage of ℓ types tends to zero as well.

[35]In some cases, merely encouraging face-to-face meetings may allow colleagues to reveal their potential interest slowly and conditionally, lessening the risk of a severe LoF from a "desperate bid". This may help explain Boudreau

This modeling of this paper can be expanded in several ways, relaxing many of our assumptions and generalizing our results. In a model allowing both inherent LoF and reputation, the effects of revealing offers on match efficiency may be complex. If a player is known to be vulnerable to LoF, his making an offer might actually be interpreted as a signal of his *confidence* that he will be accepted, thus a positive signal about his own type. Whenever a player rejects another, there is some possibility that he did so merely to avoid losing face; noting this possibility should presumably "soften the blow" to a player's reputation when he is rejected.

Relaxing the assumptions further, preferences over types may be heterogeneous, or may involve a horizontal component, this may also change the equilibrium reputation effects of revealing offers. It also may be interesting to consider the effects of a player who is either altruistic, suffering when the other player loses face, or spiteful, relishing in making others lose face. Consider a sequential game where only the first mover is vulnerable to a LoF and the second mover is a known altruist. Here the first-mover might manipulate this altruism, playing "accept" and in effect guilting the second-mover into matching with her; this could lead to inefficient matching.

Empirically, our anecdotal and referential evidence for LoF should be supplemented by experimental evidence. Field experiments (or contextual lab work) in the mold of Lee et al. (2011) will help identify preferences and beliefs. Abstract "induced values" experiments may also shed light on strategic play and coordination in our simple environment.[36] As well as strengthening the evidence for the existence of the LoF motivation, these experiments should examine the causes and correlates of LoF, and its efficiency consequences in various environments. Do people act strategically to minimize their own risk of LoF? Will they be willing to pay to preserve the anonymity of their offers? Who is most affected by loss of face and when? How can these issues be addressed to improve matching efficiency in real-world environments? We hope that future research will offer empirical answers to these questions.

Our results, and the results of the empirical project proposed above, will have important implications for government and managerial policy. Search and matching models examining the workings of labor market policies may need to adjust for the presence of LoF. Our research suggest that policies that subsidise or encourage sending applications will appear more advantageous. Organizations may want to closely consider when offers, payments, proposals, and attempts should be made transparent, and when they should be obscured. Matchmakers and middlemen in many areas, from actual marriage brokers to career "headhunters" to venture capital intermediaries may want to guarantee that unrequited offers will be kept secret. As previously noted secrecy should be helpful for the success of both international negotiations and negotiating over business mergers. Both parties may want a mutual guarantee that no offers or proposals will be leaked. Finally, we note a hub of organizations that seek to promote forms of

---

et al (2012); who found that an exogenously assigned brief face-to-face interaction between local scientists led to significant increases in their probability of collaborating.

[36]These papers involve symmetric horizontal matching preferences in an anonymous laboratory setting. Echenique and Yariv (2011) and Pais et al. (2012) allow subjects to make and reject/accept offers, one at a time for a certain duration, in small groups. Echenique and Yariv (2011) offer evidence that stability is a good predictor of market outcomes with complete information over preferences. Pais et al. (2012) find that making it costly to send offers (akin to a standard search friction) leads, unsurprisingly, to fewer and slightly less ambitious offers, and to less efficiency and less stable matchings. Incomplete information boosts both stability and efficiency. Haruvy and Utku Unver (2007) use a more rigid matching environment and find that proposer-side optimal stable matches are reached in the majority of cases regardless of the information structure. Other related papers include Kagel and Roth (2000), Nalbantian and Schotter (1995), Molis and Veszteg (2010). However, none of these (i) rely on homegrown preferences and beliefs over social interactions or partnerships, (ii) have face-to-face interaction, (iii) test the single-shot matching of our model, (iii) compare environments such as our CAE and ARE, nor (iv) have a subjects' previous choices and history reported to later matches.

cooperation and sharing that appear efficient but are not yet widely practiced. The fear of LoF may have served as a barrier to these activities in the past; in some cases setting up a "risk-free partnering exchange" may be helpful.

References

Adachi, H. (2003). A search model of two-sided matching under nontransferable utility. *Journal of Economic Theory 113*(2), 182–198.

Akerlof, G. and R. Kranton (2005). Identity and the Economics of Organizations. *Journal of Economic Perspectives 19*(1), 9–32.

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

Alaoui, L. (2010). The value of useless information. *Working Paper*.

Albrecht, J., P. Gautier, and S. Vroman (2006). Equilibrium directed search with multiple applications. *Review of Economic Studies 73*(4), 869–891.

Anderson, A. and L. Smith (2010). Dynamic matching and evolving reputations. *The Review of Economic Studies 77*(1), 3–29.

Armstrong, T. (1993). *Breaking the ice: rapprochement between East and West Germany, the United States and China, and Israel and Egypt.* United States Inst of Peace Press.

Battigalli, P. and M. Dufwenberg (2007). Guilt in games. *American Economic Review 97*(2), 170–176.

Baxter, L. and W. Wilmot (1984). Secret tests. *Human Communication Research 11*(2), 171–201.

Becker, G. (1973). A Theory of Marriage: Part I. *Journal of Political Economy 81*(4), 813.

Bonita London, Geraldine Downey, Cheryl Bonica, Iris Paltin (2007). Social causes and consequences of rejection sensitivity. *Journal of Research on Adolescence 17(3)*, 481–506.

Boudreau, K., I. Ganguli, P. Gaule, E. Guinan, and K. Lakhani (2012). Colocation and scientific collaboration: Evidence from a field experimentsns. *Working paper. Harvard Business School Electronic copy available at: http://ssrn.com/abstract=2134249.*

Bredow, C., R. Cate, T. Huston, S. Sprecher, A. Wenzel, and J. Harvey (2008). Have we met before? a conceptual model of first romantic encounters. *Handbook of relationship initiation*, 3–28.

Burdett, K. and M. Coles (1999). Long-Term Partnership Formation: Marriage and Employment. *The Economic Journal 109*(456), 307–334.

Burdett, K. and M. Coles (2006). Marriage and class. *Journal of Economic Theory 129*, 81–113.

Cawley, J. (2003). A guide (and advice) for economists on the u. s. junior academic job market. *Job Openings for Economists.*

Chade, H. (2006). Matching with noise and the acceptance curse. *Journal of Economic Theory 129*(1), 81–113.

Collins, N. and J. Goyder (2008). Speed Dating: a process of forming undergraduate student groups.

Dixit, A. (1986). Comparative statics for oligopoly. *International Economic Review 27*(1), 107–122.

Douglas, W. (1987). Affinity-testing in initial interactions. *Journal of Social and Personal Relationships 4*(1), 3–15.

Eastwick, P. and E. Finkel (2008). Speed-dating: A powerful and flexible paradigm for studying romantic relationship initiation. *Handbook of relationship initiation*, 217–234.

Eban, A. (1983). *The New Diplomacy: International Affairs in the Modern Age.* Random House NY.

Echenique, F. and L. Yariv (2011). An experimental study of decentralized matching. *Mimeo.*

Eisenberger, N. and M. Lieberman (2004). Why rejection hurts: a common neural alarm system for physical and social pain. *Trends in Cognitive Sciences 8*(7), 294–300.

Gale, D. and L. S. Shapley (1962, jan). College admissions and the stability of marriage. *The American Mathematical Monthly 69*(1), 9–15.

Geanakoplos, J., D. Pearce, and E. Stacchetti (1989). Psychological Games and Sequential Rationality. *Games and Economic Behavior 1*(1), 60–79.

Goffman, E. (2005). *Interaction Ritual: Essays in Face to Face Behavior.* Aldine Transaction.

Hahn, F. (1962). The stability of the cournot oligopoly solution. *The Review of Economic Studies*, 329–331.

Haruvy, E. and M. Utku Unver (2007). Equilibrium selection and the role of information in repeated matching markets. *Economics Letters 94*(2), 284–289.

Hitsch, G., A. Hortacsu, and D. Ariely (2010). Matching and sorting in online dating. *The American Economic Review 100*(1), 130–163.

Jonsson, C. and K. Aggestam (2008). Diplomacy and conflict resolution. *The SAGE handbook of conflict resolution*, 33.

Kagel, J. and A. Roth (2000). The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment. *The Quarterly Journal of Economics 115*(1), 201–235.

Koszegi, B. (2006). Ego Utility, Overconfidence, and Task Choice. *Journal of the European Economic Association 4*(4), 673–707.

Lee, S., M. Niederle, H. Kim, and W. Kim (2011). Propose with a rose? signaling in internet dating markets. Technical report, National Bureau of Economic Research.

MarketData Enterprises, I. (2012). *The US Dating Services Markets.* Tampa, Fla.: Marketdata Enterprises.

Mendoza-Denton, R., G. Downey, V. Purdie, A. Davis, and J. Pietrzak (2002). Sensitivity to status-based rejection: implications for african american students' college experience. *Journal of personality and social psychology 83*(4), 896.

Milgrom, P. and C. Shannon (1994). Monotone comparative statics. *Econometrica: Journal of the Econometric Society*, 157–180.

Molis, E. and R. Veszteg (2010). Experimental results on the roommate problem. *CORE DP 11.*

Moreno, A. (2008). Group Fairness and Game Theory. *School of Economics Working Papers.*

Nalbantian, H. and A. Schotter (1995). Matching and efficiency in the baseball free-agent system: an experimental examination. *Journal of Labor Economics*, 1–31.

Pais, J., A. Pinter, and R. Veszteg (2012). Decentralized matching markets: a laboratory experiment.

Putnam, R. (2001). *Bowling alone.* Simon & Schuster.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *American Economic Review 83*(5), 1281–1302.

Risemberg, R. (1999, July 6). Date matching methods. US Patent 5,920,845.

Rosenfeld, M. and R. Thomas (2012). Searching for a mate: The rise of the internet as a social intermediary.

Shanteau, J. and G. Nagy (1979). Probability of acceptance in dating choice. *Journal of Personality and Social Psychology 37*(4), 522.

Vives, X. (2005). Complementarities and games: New developments. *Journal of Economic Literature*, 437–479.

Williams, K. (2007). Ostracism. *Annu Rev Psychol 58*, 425–52.

A Terminal information sets and game trees

The set of end nodes of the game, defined by their histories, is $h = \{h1, h2, h3, h4\} = \{AA, AR, RA, RR\}$.[37] Let $\bar{I}_f$ be the collection of f's terminal information sets over these endnodes, and $\bar{I}_m$ be m's information partition. Since neither player "has the move" at the terminal node, we give each history two boxes to depict each player's terminal information set; $h_j(m)$ and $h_j(f)$ are the same (for $j \in \{1, 2, 3, 4\}$).

Two versions of this game are illustrated in the trees below, which display only *material* payoffs, followed by a specification of the terminal information partitions for each case.
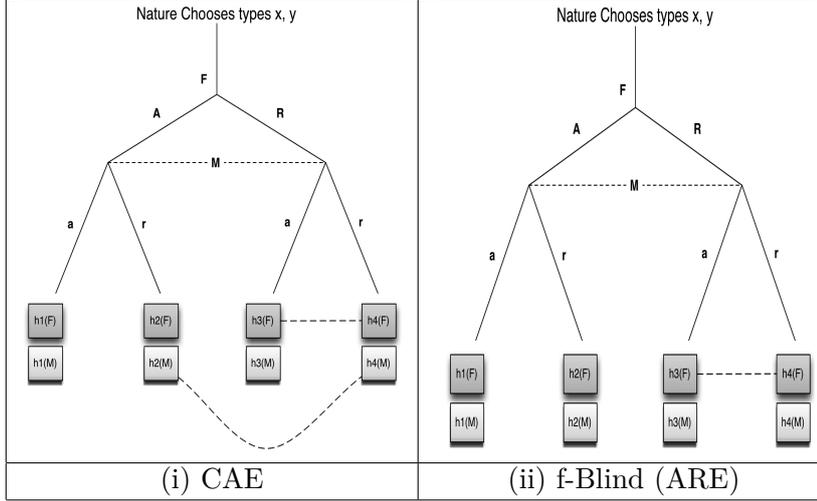

Figure 3: Terminal Information structures
(i) Conditionally Anonymous (CAE): $\bar{I}_F = \{\{AA\}, \{AR\}, \{RA, RR\}\}$ and $\bar{I}_m = \{\{AA\}, \{AR, RR\}, \{RA\}\}$.
(ii) f is "blind": $\bar{I}_f = \{\{AA\}, \{AR\}, \{RA, RR\}\}$; $\bar{I}_M = \{\{AA\}, \{AR\}, \{RA\}, \{RR\}\}$.

Note, for example that in the simultaneous CAE the $\{AA\}$ terminal information set is a singleton for both players, while the histories where a player played "reject" are part of the same terminal information set (for that player).

B Proofs

Exhaustive proof that players without LoF use threshold strategies in settings with and without LoF

Define $q(x, x')$ as the probability that a player of type x accepts a match of type x'. Writing this in terms of likelihood ratios, the acceptance condition becomes $\frac{p f_h(s) q(h,h) \tilde{h} + (1-p) f_\ell(s) q(\ell,h) \tilde{\ell}}{p f_h(s) q(h,h) + (1-p) f_\ell(s_f) q(\ell,h)} \geq \delta \tilde{h}$, which simplifies to

$$\frac{f_h(s)}{f_\ell(s)} pq(h, h)(\tilde{h} - \delta\tilde{h}) \geq (1 - p)q(\ell, h)(\delta\tilde{h} - \tilde{\ell}). \tag{20}$$

Using the monotone likelihood ratio property and our other assumptions over the density we know the ratio $\frac{f_h(s_f)}{f_\ell(s_f)}$ increases from 0 to 1 over the range of s. The right hand side is positive as

---

[37]We leave nature's move out of these histories; it does not affect our discussion. For completeness we can assume that players never learn the other players' types. Thus, in our model LoF will only depend on the conditional expectation of the other player's type, not the type itself.

long as ℓ's accept with at least some positive probability. We can rule out an equilibrium where ℓ's always reject even when ℓ's suffer from LoF. Regardless of h's behavior, since ℓ's will always reciprocate an acceptance, if an ℓ observes a low enough signal the expected benefit of marrying an ℓ will outweigh the expected LoF from being rejected by an h. Ruling this out implies that h's must reject when they observe the lowest signals, as these almost surely come from ℓ's. Thus h's must have a cutoff best response $\hat{s} > \underline{s}$.

Proof of Lemma 2

The use of thresholds $\hat{s}_h, \hat{s}_\ell \in [\underline{s}, \overline{s}]$ follows from condition (12) and the monotone likelihood property, see text. Since $\delta\tilde{h} > \tilde{\ell} > \delta\tilde{\ell}$ by assumption, male h types accept if $s_f \geq \hat{s}_h$ and male ℓ types accept if $s_f \leq \hat{s}_\ell$. $\hat{s}_h > \underline{s}$ and $\hat{s}_\ell > \underline{s}$, since $\delta\tilde{h} > \tilde{\ell} > (1-\delta)\tilde{\ell}$ and for $s_f = \underline{s}$ the probability that the female match is of type ℓ is 1. Noting that condition (13) implies condition (12) for $\tilde{x} = \tilde{\ell}$ and that condition (14) implies that (12) does not hold for $\tilde{x} = \tilde{h}$ establishes the lemma.

Proof of Proposition 1

The changes in $\hat{s}_f$ and $\hat{s}_h$ from $\hat{s}$ associated with the ϵ increase in L are determined by the system of equations (18) and (15). The total differential of (18) for $L = \epsilon$ small enough for the ceiling $\hat{s}_\ell^* = \overline{s}$ to hold is

$$\frac{\partial \frac{f_h(\hat{s}_f)}{f_\ell(\hat{s}_f)}}{\partial s} \frac{p}{1-p} \frac{(1-\delta)\tilde{h}}{\delta\tilde{h} - \tilde{\ell}} \Delta\hat{s}_f = \frac{f_h(\hat{s}_h)}{(1 - F_h(\hat{s}_h))^2} \Delta\hat{s}_h,$$

where the Δ terms denote the changes in $\hat{s}_f$ and $\hat{s}_h$. Note that this already implies that the equilibrium cutoffs for high females and males move in the same direction as long as L is sufficiently small to ensure that the ceiling $\hat{s}_\ell = \overline{s}$. For marginal changes we have

$$\frac{\partial \hat{s}_f}{\partial \hat{s}_h} = \frac{\frac{f_h(\hat{s}_h)}{(1 - F_h(\hat{s}_h))^2}}{\frac{\partial \frac{f_h(\hat{s}_f)}{f_\ell(\hat{s}_f)}}{\partial s} \frac{p}{1-p} \frac{(1-\delta)\tilde{h}}{\delta\tilde{h} - \tilde{\ell}}} > 0. \tag{21}$$

Conducting a similar exercise for (15) yields

$$\frac{\partial \frac{f_h(\hat{s}_h)}{f_\ell(\hat{s}_h)}}{\partial s} \frac{p}{1-p} \frac{(1-\delta)\tilde{h}}{\delta\tilde{h} - \tilde{\ell}} \Delta\hat{s}_h = \frac{f_h(\hat{s}_f)(L + (1-\delta)\tilde{h})\Delta\hat{s}_f + F_h(\hat{s}_f)\delta\tilde{\ell}}{(1-\delta)\tilde{h}(1 - F_h(\hat{s}_f)(1 + L/((1-\delta)\tilde{h})))^2}. \tag{22}$$

Plugging in $\Delta\hat{s}_f$ from above and focusing on the neighborhood of $L = 0$, where $\hat{s}_f = \hat{s}_h = \hat{s}^*$, yields that for marginal variations in L

$$\frac{\partial \hat{s}_f}{\partial L} = \frac{f_h(\hat{s}^*)\frac{F_h(\hat{s}^*)}{(1-\delta)\tilde{h}}}{(1 - F_h(\hat{s}^*))^4 \left(\frac{\partial \frac{f_h(\hat{s}^*)}{f_\ell(\hat{s}^*)}}{\partial s}\right)^2 \left(\frac{p}{1-p}\frac{(1-\delta)\tilde{h}}{\delta\tilde{h} - \tilde{\ell}}\right)^2 - f_h(\hat{s}^*)^2}.$$

That is, a marginal increase of L at $L = 0$ yields a (weak) increase of $\hat{s}_f$ (and $\hat{s}_h$ by (21)) if, and only if

$$\frac{\partial \hat{s}_f}{\partial \hat{s}_h} = \frac{f_h(\hat{s}^*)}{(1 - F_h(\hat{s}^*))^2} \frac{p}{1-p} \frac{(1-\delta)\tilde{h}}{\delta\tilde{h} - \tilde{\ell}} \frac{\partial \frac{f_h(\hat{s}^*)}{f_\ell(\hat{s}^*)}}{\partial s} < 1.$$

Proof of Lemma 4

Trembling hand perfection requires the equilibria of a sequence of games with ϵ trembles to converge to the equilibrium of the game without trembles as ϵ approaches zero. In a game perturbed by an ϵ tremble agents play mixed strategies that place at least probability ϵ on each pure strategy. In the unperturbed game with $L > 0$ males strictly prefer to reject with certainty if female h types do so as well. Hence, also for small perturbations resulting in $q_m^{\epsilon}(h, h) > 0$ (but still small) males strictly prefer to reject with certainty. Hence, male h type's best response converges to its counterpart in the equilibrium in the unperturbed game as ϵ converges to 0. This implies that also female h types' best reply converges to its counterpart in the equilibrium in the unperturbed game. Hence, the equilibrium is indeed trembling hand perfect.

The last statement in the proposition follows from the fact that $\hat{s}_m = \overline{s}$, which implies that condition (13) cannot hold, thus Lemma 2 must apply.

The proof for tatonnement stability follows immediately from the above.

Details for corollary to lemma 4

(i) In the benchmark case all ℓ accept and some h do. In the CFE with LoF no h accept, but all male ℓ accept. Suppose also the "best case" that all female ℓ's also accept. (As noted, this is counterfactual; female ℓ's will reject against the highest signals). This yields the reduced marriage rates given in the corollary.

(ii) The reduction in aggregate payoffs follows directly from the linear marriage production function.

(iii) The argument for the reduced expected surplus of each gender/type relies on a revealed preference argument. Consider going in the opposite direction, from CFE to the benchmark. The strict increase in the rate of acceptance for h's requires that h's must be better off when they accept, as their outside option remains the same, and are thus better off in expectation. (An analogous argument holds for the increase in the acceptance rate for female ℓ's shown later). Furthermore, when playing A, both ℓ's and h's are better off when h's are in the market; their expected mate will be of higher quality.

As argued earlier, even without a primal loss of face, being observed playing *accept* may have a negative reputational consequence (which itself may be painful). In a multi-period interaction this may also lead to *material* losses, including worse matching prospects in later periods. We demonstrate this below in a two-period model. We compare the case where a player's choice in the first period is private, to a case where it is revealed to her second-period match. We show that in the latter case playing *accept* in the first period worsens one's continuation value, making first-period incentives equivalent to our primal LoF model in Section 4.

B.1 A Repeated Matching Market

Basic payoffs and the information structure follows from section 3.5. However, we now assume that the market remains open for two periods. We give the timing below; note that stages 1-3 are essentially the same as above.

We focus on settings that result in stable, interior equilibria, i.e., settings where Condition 7 holds.

Timing

1. Individuals in M and F are matched to each other randomly. In each match (m, f) each individual i obtains a noisy signal s_j about the other one's type x_j.

2. After observing the signal individuals simultaneously decide on whether to accept or reject the match.

3. Pairs in which both individuals played *accept* form a marriage and are removed from the market and receive payoffs.

4. The remaining individuals are again matched into pairs (m, f) randomly. Again, in each match (m, f) each individual obtains a noisy signal s_j about the other's type x_j.[38]

5. *In an "Asymmetric Partial Revelation Environment" (APRE), but not the "Conditionally Anonymous Environment" (CAE), females observe their (male) match's action in the previous stage, but not vice-versa.*

6. After observing the signal individuals simultaneously decide on whether to accept or reject the match.

7. Pairs in which both individuals accepted form a marriage, all others remain single. The market closes and payoffs are realized.

B.2 Conditionally anonymous environment (CAE)

We solve backwards from the second stage; this solution is identical to that of the one-period model without LoF, re-labeling some variables; we thus give only a few key equations and results below. Denote the share of h agents on each side in period 2 by p_2, which will depend on players' equilibrium strategies. From the assumption that $f_h(\underline{s}) = 0$ we know that a strictly positive measure of agents remains in the market at $t = 2$.[39] Thus $0 < p_2 < 1$. In the second period, an agent who receives a signal s_j assesses the probability of facing a high type as

$$pr(h|s_j) = \frac{p_2 f_h(s_j)}{p_2 f_h(s_j) + (1 - p_2) f_\ell(s_j)}.$$

B.2.1 Equilibrium Behavior

CAE: Second period

As there is no further reputation motive in the final period, nor an intrinsic LoF, nor is previous play observed in the CAE, this period is equivalent to the CAE (or benchmark model without

---

[38] For brevity, we do not separately index the first and second period match, nor the first and second period signals – these will be clear from context.

[39] To violate this both types would have to play "always accept" in the first stage, expecting to remain unmarried with probability 1 conditional on reaching stage 2. This would be an equilibrium if the expected match quality in period 1 were high enough for an h type to accept regardless of the signal:

$$p \geq \frac{f_\ell(\underline{s})(\delta \tilde{h} - \tilde{\ell})}{f_\ell(\underline{s})(\delta \tilde{h} - \tilde{\ell}) + f_h(\underline{s})(1 - \delta)\tilde{h}}.$$

This case is excluded by the assumption that $f_h(\underline{s}) = 0$. Moreover, if this assumption were not made and this were an equilibrium, it would not be stable nor trembling hand perfect, since an agent facing a signal $s = \underline{s}$ in $t = 1$ would prefer to take another chance in $t = 2$ if the proportion of ℓ and h agents were to remain the same, as it would under a random tremble.

LoF) from section 3.5, replacing p with p_2, and noting that s refers to the second-period signal. The results follow (restating the earlier lemma and corollary with minor adjustments to notation).

Lemma 5. *[CAE Second Stage] In a CAE, if an interior stable equilibrium exists (conditions for this are as in the main text), then for the subgame starting in period 2:*

1. *low types always accept,*

2. *high types use symmetric cutoff strategies, accepting if $s > \hat{s}_m = \hat{s}_f := \hat{s}^*$, and*

3. *$\hat{s}^* \in (\underline{s}, \overline{s})$.*

Note $\hat{s}^*$ is defined by equation (5), replacing p with p_2.

Corollary. *Expected payoffs for types ℓ and h in an interior stable equilibrium of the second-period subgame of the CAE (or for any strategy profile where low types always accept) are*

$$v_2(\ell) = \delta\tilde{\ell} + p_2(1 - F_\ell(\hat{s}^*))(\tilde{h} - \delta\tilde{\ell}) + (1 - p_2)(\tilde{\ell} - \delta\tilde{\ell}), and$$
$$v_2(h) = \delta\tilde{h} + p_2(1 - F_h(\hat{s}^*))^2(\tilde{h} - \delta\tilde{h}) - (1 - p_2)(1 - F_\ell(\hat{s}^*))(\delta\tilde{h} - \tilde{\ell}). \tag{23}$$

Note $v_2(h) > v_2(\ell)$.[40] *The number of marriages is $(1 - p_2)^2 + 2p_2(1 - p_2)(1 - F_\ell(\hat{s}^*)) + p_2^2(1 - F_h(\hat{s}^*))^2$, which strictly decreases in $\hat{s}^*$.*

CAE: First Stage

A similar reasoning applies to the first stage: an agent of type x_i who observes a signal s_j finds it profitable to accept if and only if the expected payoff exceeds her continuation value $v_2(x_i)$ after playing reject. As, in the CAE, the continuation value from rejecting or being rejected are both zero, we only need to see if the payoff from an expected marriage exceeds this. Stating this in terms of gains relative to the continuation value we have:

$$p f_h(s) q_1(x, h)(\tilde{h} - v_2(x)) + (1 - p) f_\ell(s) q_1(x, \ell)(\tilde{\ell} - v_2(x)) \geq 0. \tag{24}$$

Recall, a high type prefers marrying a high type over remaining single, which she prefers over marrying a low type; thus her continuation value satisfies $\tilde{\ell} < \delta\tilde{h} < v_2(h) < \tilde{h}$ (the strictness of the latter inequalities follows from $0 < p_2 < 1$, previously demonstrated, and from an interior equilibrium in period 2, under conditions specified). Thus, for a high type the first additive term in inequality (24) must be positive and the second term negative. By the monotone likelihood property, as the signal s increases the first term increases in magnitude and the second term decreases in magnitude (and this holds for both types). Thus the high type will set a floor threshold. By a similar argument, for a low type ($x = \ell$) the first term must be positive but the second term may have either sign. If it is negative the low type must also set a floor threshold. If both terms are positive the low type will always accept in the first period (a trivial floor at $\underline{s}$).

---

[40]This must hold as high types could always adopt the same "always accept" strategy as low types and gain a strictly higher payoff than low types. In doing so, both high and low types would always marry when they meet a low type, but high types would be more likely to marry when meeting another high type (and high type's unmarried payoff are also higher).

Thus both types will set a floor threshold: there are values $\hat{s}_x \in [\underline{s}, \overline{s}]$ for $x = \ell, h$, such that the agent accepts only if $s \geq \hat{s}_x$. Similarly to the second stage, where interior, these are implicitly defined by

$$p f_h(\hat{s}_x)(1 - F_x(\hat{s}_h))(\tilde{h} - v_2(x)) = (1 - p) f_\ell(\hat{s}_x)(1 - F_x(\hat{s}_\ell))(v_2(x) - \tilde{\ell}), \tag{25}$$

where the left side represents the expected benefit from marrying high relative to staying alone, and the right side the expected "loss" from marrying low relative to staying alone.

Lemma 6. *The first period cutoffs satisfy $\hat{s}_h > \hat{s}_\ell$.*

Proof: Rearranging (25) yields cutoff values $\hat{s}_h$ and $\hat{s}_\ell$ defined by

$$\frac{p f_h(\hat{s}_h)}{(1-p) f_\ell(\hat{s}_h)} = \frac{v_2(h) - \tilde{\ell}}{\tilde{h} - v_2(h)} \frac{1 - F_h(\hat{s}_\ell)}{1 - F_h(\hat{s}_h)} \text{ and}$$

$$\frac{p f_h(\hat{s}_\ell)}{(1-p) f_\ell(\hat{s}_\ell)} = \frac{v_2(\ell) - \tilde{\ell}}{\tilde{h} - v_2(\ell)} \frac{1 - F_\ell(\hat{s}_\ell)}{1 - F_\ell(\hat{s}_h)}.$$

Since $v_2(h) > v_2(\ell)$ as shown above, the contradiction to the lemma, $\hat{s}_h \leq \hat{s}_\ell$ implies that

$$\frac{1 - F_h(\hat{s}_\ell)}{1 - F_h(\hat{s}_h)} \leq \frac{1 - F_\ell(\hat{s}_\ell)}{1 - F_\ell(\hat{s}_h)}.$$

Rewriting this in terms of integrals yields

$$1 + \frac{\int_{\hat{s}_h}^{\hat{s}_\ell} f_\ell(s) ds}{\int_{\hat{s}_\ell}^{\overline{s}} f_\ell(s) ds} \leq 1 + \frac{\int_{\hat{s}_h}^{\hat{s}_\ell} f_h(s) ds}{\int_{\hat{s}_\ell}^{\overline{s}} f_h(s) ds}.$$

This becomes

$$\frac{\int_{\hat{s}_h}^{\hat{s}_\ell} f_h(s) ds}{\int_{\hat{s}_h}^{\hat{s}_\ell} f_\ell(s) ds} \geq \frac{\int_{\hat{s}_\ell}^{\overline{s}} f_h(s) ds}{\int_{\hat{s}_\ell}^{\overline{s}} f_\ell(s) ds}.$$

The monotone likelihood ratio property implies the contrary, since

$$\frac{f_h(s)}{f_\ell(s)} < \frac{f_h(\hat{s}_\ell)}{f_\ell(\hat{s}_\ell)} < \frac{f_h(s')}{f_\ell(s')},$$

for all $s < \hat{s}_\ell < s'$.

Thus high types will be more selective than low types in stage 1. □

Equation (25) leads to several results. Recalling assumption 2, that extreme signals fully reveal types, low types will "accept always" only if $v_2(\ell) \leq \tilde{\ell}$. This cannot be ruled out: the low types' continuation value may be below the value of marrying low, as they may be rejected in stage 2, which would leave them worse off.

In contrast, high types do not "accept always" – this would require $v_2(h) \leq \tilde{\ell}$, but as we have shown $v_2(h) > \tilde{h}$ and $\delta\tilde{h} > \tilde{\ell}$ by assumption. High types will also not "reject always"; this would be optimal only if $v_2(h) \geq \tilde{h}$, which could hold only if there are no low types in stage 2. But we know $p_2 < 1$: even if all ℓ's accept in stage 1 some will meet high types, give off low signals, and be rejected, surviving to stage 2. Hence, $\underline{s} < \hat{s}_h < \overline{s}$, and $\underline{s} \leq \hat{s}_\ell < \hat{s}_h$. This fully characterizes equilibrium strategies, summarized in the following proposition.

Proposition 2. *[CAE Equilibrium] In an equilibrium of the repeated matching market under the CAE low types always accept in stage 2 and accept in stage 1 only if $s_j \geq \hat{s}_\ell$. $\hat{s}_\ell$ solves*

equation 25 (with $x = \ell$), and $\underline{s} \leq \hat{s}_\ell < \overline{s}$. High types accept in stage 2 if and only if $s_j \geq \hat{s}_2$ (where $\hat{s}_2$ solves the equivalent of equation 5) and accept in stage 1 if and only if $s_j \geq \hat{s}_h$, where $\hat{s}_h$ solves equation 25 (with $x = h$). $\hat{s}_h \in (\underline{s}, \overline{s})$ and $\hat{s}_h > \hat{s}_\ell$.

B.3 Asymmetric partial revelation environment

We next consider an environment analogous to the ARE in the main text. In an Asymmetric Partial Revelation Environment (APRE) a male's stage 1 choice will be observed if he is present in stage 2. If we assume, conforming to intuition, that in stage one high types are at least as selective as low types, then the stage 2 reputational consequences of playing *Accept* will lead to a loss of continuation value similar to the intrinsic LoF.[41] (However, here the loss of continuation value here may depend on one's type; we discuss this below.)

Thus we now suppose that in stage 2 a female i matched to a male agent j not only observes a signal s but also j's stage 1 action $A_j \in \{A; R\}$. Hence a male is characterized by pairs $xA \in \{\ell A; \ell R; hA; hR\}$, which we will refer to as an "attribute." Denote the measure of a male with each attribute by p^m_{xA}. Then $p^m_2 = p^m_{hA} + p^m_{hR}$ and as above $0 < p^m_2 < 1$.

Suppose a female agent i observes signal s_j and past action A_j. Conditionally on being accepted and on the other observables, a female i assesses the probability that her match is type h with

$$Pr_f(h|s, x_i, A_j, acc) = \frac{f_h(s)p_{hA_j}q^f_2(x_i, A_j, h)}{f_h(s)q^f_2(x_i, A_j, h)p_{hA_j} + f_\ell(s)(1 - p_{hA_j})q^f_2(acc|x_i, A_j, \ell)},$$

where $q^f_2(x_i, A_j, x_j)$ is the probability that in stage 2 a female of type x_i is accepted by a male of type x_j who played action A_j in stage 1.

A male j will assess the probability that, given he is accepted and his prior action, his match is type h with

$$Pr_m(h|s, x_j, A_j, acc) = \frac{f_h(s)p_2q^m_2(x_j, A_j, h)}{f_h(s)p_2q^m_2(x_j, A_j, h) + f_\ell(s)(1 - p_2)q^m_2(x_j, A_j, \ell)},$$

where $q^f_2(x_i, A_j, x_j)$ is the probability that in stage 2 a male of type x_j who played action A_j in stage 1 is accepted by a female of type x_i.

As usual, low-type agents of both genders will always find a marriage profitable regardless of the signal they observe and the previous action of their match, since $u_x(\ell) > \tilde{\ell}$. Thus, for both genders, in stage 2 low types always accept. A high type agent finds accepting profitable (as before) if and only if, conditional on being accepted, the expected marriage is a favorable one. I.e., for a high-type female i, if

$$\frac{p_{hA_j}f_h(s)q^f_2(h, A_j, h)\tilde{h} + (1 - p_{hA_j})f_\ell(s)q^f_2(h, A_j, \ell)\tilde{\ell}}{p_{hA_j}f_h(s)q^f_2(h, A_j, h) + (1 - p_{hA_j})f_\ell(s)q^f_2(h, A_j, \ell)} > \tilde{h},$$

and for a high-type male j, if

$$\frac{p_2f_h(s)q^m_2(h, A_j, h)\tilde{h} + (1 - p_2)f_\ell(s)q^m_2(h, A_j, \ell)\tilde{\ell}}{p_2f_h(s)q^m_2(h, A_j, h) + (1 - p_2)f_\ell(s)q^m_2(h, A_j, \ell)} > \tilde{h}.$$

---

[41]We show the existence of this more intuitive equilibrium, but we do not rule out other equilibria. We save this for later work more focused on modelling reputation.

Since the acceptance probabilities of one's match do not depend on the realization of the signal s one observes, the monotone likelihood ratio property implies that the left-hand side strictly increases in the observed signal s and a high-type agent uses a threshold strategy of the type "accept if and only if $s \geq \hat{s}_2$" as above. Let $\hat{s}_m(h, A_j)$ denote the cutoff for a high-type male who played A_j in stage 1, and $\hat{s}_f(h, A_j)$ denote the cutoff for a high-type female (the low types always accept) matched to a male who played A_j.

Using the conditions above the threshold values are thus implicitly defined by

$$\frac{p_{hA_j} f_h(\hat{s}_f(A_j))(1 - F_h(\hat{s}_m(A_j)))\tilde{h} + p_{\ell A_j} f_\ell(\hat{s}_f(A_j))\tilde{\ell}}{p_{hA_j} f_h(\hat{s}_f(A_j))(1 - F_h(\hat{s}_m(A_j))) + p_{\ell A_j} f_\ell(\hat{s}_f(A_j))} = \delta\tilde{h}, \tag{26}$$

for females, and by

$$\frac{p_2 f_h(\hat{s}_m(A_j))(1 - F_h(\hat{s}_f(A_j)))\tilde{h} + (1 - p_2) f_\ell(\hat{s}_m(A_j))\tilde{\ell}}{p_2 f_h(\hat{s}_m(A_j))(1 - F_h(\hat{s}_f(A_j))) + (1 - p_2) f_\ell(\hat{s}_m(A_j))} = \delta\tilde{h} \tag{27}$$

for males. This yields four different cutoff values: $\hat{s}_f(A)$ for a female when facing a male who accepted in stage 1, $\hat{s}_f(R)$ for female facing a male who rejected, $\hat{s}_m(A)$ for a male who accepted in stage 1 and $\hat{s}_m(R)$ for a male who rejected.

Note that, in contrast to the CAE, we must consider cases where types are known for certain, conditional on previous acceptance behavior. This yields $\hat{s}_f(A_j) = \underline{s}$ if $Pr(h|A_j) = 1$ and $\hat{s}_f^{A_j} = \overline{s}$ if $Pr(h|A_j) = 0$. I.e., there is a possibility of a fully separating equilibrium where h and ℓ play different actions in the first stage allowing perfect revelation of type in the second stage. It is, however, straightforward to bring this possibility to a contradiction with equilibrium play.[42]

The cutoff values defined in equations 26 and 27 satisfy

$$\frac{p_{hA}}{p_{\ell A}} \frac{f_h(\hat{s}_f(A))}{f_\ell(\hat{s}_f(A))}(1 - F_h(\hat{s}_m(A))) = \frac{\delta\tilde{h} - \tilde{\ell}}{(1 - \delta)\tilde{h}} = \frac{p_{hR}}{p_{\ell R}} \frac{f_h(\hat{s}_f(R))}{f_\ell(\hat{s}_f(R))}(1 - F_h(\hat{s}_m(R)))$$

$$\text{and } \frac{f_h(\hat{s}_m(A))}{f_\ell(\hat{s}_m(A))}(1 - F_h(\hat{s}_f(A))) = \frac{\delta\tilde{h} - \tilde{\ell}}{(1 - \delta)\tilde{h}} = \frac{f_h(\hat{s}_m(R))}{f_\ell(\hat{s}_m(R))}(1 - F_h(\hat{s}_f(R))).$$

The second statement in turn implies that $\hat{s}_f(A) > \hat{s}_f(R) \Leftrightarrow \hat{s}_m(A) > \hat{s}_m(R)$, i.e., if high type females are more choosy with respect to males who chose a certain action, males who chose that action will be more choosy than males who did not. Suppose that $p_{hA}/p_{\ell A} < p_{hR}(p_{\ell R}$. Then with condition 7 from above, ensuring an interior equilibrium exists, $\hat{s}_f(A) \leq \hat{s}_f(R)$ yields a contradiction. Hence, $\hat{s}_f(A) > \hat{s}_f(R)$ if $p_{hA}/p_{\ell A} < p_{hR}/p_{\ell R}$.

To verify that this screening behavior may indeed lead to a reputational loss of face, motivating our setup above, note that expected stage 2 payoffs for the different types of male players are

$$v_2(\ell A_i) = \delta\tilde{\ell} + (1 - p_2)(1 - \delta)\tilde{\ell} + p_2(1 - F_\ell(\hat{s}_f(A_i)))(\tilde{h} - \delta\tilde{\ell}) \text{ and}$$

$$v_2(h A_i) = \delta\tilde{h} + (1 - p_2)(1 - F_\ell(\hat{s}_m(A_i))(\tilde{\ell} - \delta\tilde{h}) + p_2(1 - F_h(\hat{s}_m(A_i))(1 - F_h(\hat{s}_f(A_i)))(1 - \delta)\tilde{h}.$$

---

[42]Two possibilities emerge: h's play "always accept" and ℓ's "always reject" in stage 1, in which case an ℓ would have a profitable deviation by playing 'always accept" in stage 1. Second, h's play "always reject" and ℓ "always accept" in stage 1, in which case an ℓ could do better by playing "always reject" in stage 1, which gives players of both genders (through imitating the good signal for male, and staying in the game to access a better pool for females) a positive probability to marry an h type in stage 2, and does not decrease the probability of being accepted by an ℓ type.

This means that low males will face a reputational loss of face (i.e., $L = v_2(\ell A) - v_2(\ell R) > 0$) if $\hat{s}_f(A) > \hat{s}_f(R)$, i.e., females are pickier when facing a male who accepted in round 1. Note that also high type males will face a reputational loss of face (i.e., $v_2(hA) > v_2(hR)$) if $\hat{s}_f(A) > \hat{s}_f(R)$, using an envelope argument, noting that the effect of marginal changes of $\hat{s}_x(.)$ will satisfy 27.

That is, $L > 0$ if $\hat{s}_f(A) > \hat{s}_f(R)$. From above we know that for any $L \leq 0$ high type males are more selective than low males, so that $p_{hA} < p_{hR}$ and $p_{\ell R} < p_{\ell A}$, which implies that indeed $\hat{s}_f(A) > \hat{s}_f(R)$. The following proposition summarizes these derivations.

Proposition 3. *In an APRE in stage 2 a male player who has played accept in stage 1 has lower expected payoff than if he had played reject. That is, there is reputational loss of face in the two period matching market.*