

Collegio Carlo Alberto



Reference Dependence, Risky Projects and
Credible Information Transmission

Edoardo Grillo

No. 331

December 2013

Carlo Alberto Notebooks

www.carloalberto.org/research/working-papers

Reference Dependence, Risky Projects and Credible Information Transmission*

Edoardo Grillo[†]

Collegio Carlo Alberto

This draft: March 2013

Abstract

We consider a model in which an informed Sender (S) makes an announcement concerning the quality of a project that an uninformed Receiver (R) can undertake or not. We study the role that reference dependence and loss aversion may play in affecting S' communication strategy and we show that they may give rise to credible information transmission. This happens because in our model inaccurate information has two effects: it leads R to choose the action S prefers in the short run, but it also generates unrealistic expectations that, through their effect on reference point's formation, may induce R to take actions that hurt S in the long run. On the contrary, credible information transmission is not possible if R is a standard expected utility maximizer (no reference dependence) or if his loss aversion is either too high or too low.

JEL CODES: D03, D81, D83

*I am indebted to my adviser, Stephen Morris, for his encouragement throughout the project. I am also grateful to Roland Bénabou and Wolfgang Pesendorfer for their suggestions and comments. I also want to thank Dilip Abreu, Sylvain Chassang, Faruk Gul, Juan Ortner and Takuo Sugaya and the participants in seminars at Collegio Carlo Alberto, IMT Lucca, McGill University, Paris HEC, Pompeu Fabra, Princeton University, University of Exeter Business School. I am grateful to Unicredit&Universities Foundation for its financial support. Needless to say, I take full responsibility for all remaining errors.

[†]Collegio Carlo Alberto, Via Real Collegio 30, Moncalieri (Turin), Italy. Email: edoardo.grillo@carloalberto.org).

1 Introduction

Strategic information transmission has received considerable attention by the economic literature.¹ In its simplest formulation, this literature focuses on settings in which an informed Sender has some private information about the state of the world and tries to transmit it to an uninformed Receiver who may condition his own behavior on the message received. In standard settings, communication may affect Receiver's behavior insofar it modifies his beliefs concerning the state of the world. In this paper we show that the change in beliefs induced by the Sender's announcements may also modify the Receiver's reference point and we use this insight to prove that reference dependence and loss aversion may provide credibility to the Sender's announcements. This result is obtained in a model where the Receiver has to decide whether to participate in a project whose probability of success is known only to the Sender; in such a setting, if the Receiver were a standard expected utility maximizer, credible communication would not be possible due to the conflict of interests between the two parties.

To get an intuition, consider the following strategic interaction. S and R (respectively, she and he) are two researchers who are deciding whether to start a coauthorship. The joint effort would generate one of two types of manuscripts: high-potential manuscripts have high probability to appear in a highly ranked general-interest journal, while standard ones are published in such journals only with lower probability.² All manuscripts that do not appear in a top general-interest journal (either because they are not submitted in the first place or because they get rejected) are published in a field journal. Suppose the project belongs to the field on which S is actively doing research, so that she knows the literature and she can correctly assess the quality of the project. On the contrary, R is not an expert of this particular field and is uncertain about the project's quality, but his contribution is needed to write the manuscript (for instance, he may have technical skills that S does not possess). The differences in research interests translate into different evaluations of journals: although both researchers aim at a publication in a general interest journal, S also gets a positive utility from a field journal one, while R does not. As a result, whereas S would like to undertake both types of projects, R would like to participate only if the project has high likelihood of appearing in a general interest journal. Further suppose that, *conditional on having drafted the manuscript*, (i) R may learn the actual quality of the project (for instance, he may get

¹Crawford and Sobel, 1982 analyze the benchmark case of "cheap" (that is, costless and not verifiable) communication between an informed Sender and an uninformed Receiver. In particular, this paper focuses on a static game with one round of communication, one Sender and one Receiver. This model has been subsequently generalized in many directions (an incomplete review of this extensive literature on includes Aumann and Hart, 2003, Battaglini, 2002, Farrell and Gibbons, 1989, Green and Stokey, 2007, Krishna and Morgan, 2004). The cheap talk benchmark has also been used to study players' communication about their intentions concerning a subsequent strategic interaction (see, for instance Aumann, 1990, Baliga and Morris, 2002, Charness, 2000, Farrell, 1988, 1995, Rabin, 1994). Departing from the assumption of "cheap talk", this literature also analyzes communication costs (Kartik et al., 2007, Kartik, 2009) or verifiability of the announcements made (Dziuda, 2011, Ottaviani and Sorensen, 2006, Seidmann and Winter, 1997).

²These probabilities may capture some randomness in the drafting of the manuscript or in the refereeing process that leads to publications.

feedback during some conferences or receive detailed comments from other researchers), and (ii) both researchers would agree that manuscripts should be submitted to a top general interest journal only if the probability of publication is high.

Notice that both S and R would like to share the information that the project has high probability of being published in a top journal, so that R could confidently start working on it. Nevertheless, under standard utilities, this piece of information cannot be transmitted. The reason is as follows. S knows that if R learns the true quality of the project, he will insist on a submission to a highly ranked general interest journal only if the probability of publication is high. Thus, even if this probability is low, she will overstate its quality in the attempt to induce R's participation and secure, at least, a field-journal publication. Since R foresees this happening, he will not trust S' message and he will prefer not to start the project. Therefore, even high-potential projects will not be undertaken.

Now assume R has reference-dependent preferences; thus, his *total utility* depends not only on the *consumption utility* we described before, but also on the comparison between this utility and a *reference utility*: any positive (negative) deviation from the reference utility is associated with a gain (loss). Furthermore, assume R is loss averse, namely he dislikes losses more than he likes equal-size gains. Under these assumptions, if R started the coauthorship believing that the project has high probability of being published in a top general interest journal (for instance, because S said so), he may oppose to a submission to a field journal even if he realizes that his initial expectations were wrong. Intuitively, reference dependence and loss aversion may lead R to run after the slim probability of getting a top publication in order to avoid the loss that he would experience otherwise. This type of behavior may, in turn, provide S with the credibility she needs. Indeed, since lying can now lead to inefficient submissions to highly ranked journals, it is no longer costless from the Sender's point of view and she may prefer to tell the truth. As a result S may be able to induce R's participation on high-potential projects.

In this paper, we formalize this mechanism. In particular, we consider a model where the Receiver exhibits reference dependence à la Köszegi and Rabin and has to decide whether to join a project whose probability of success can be either low or high. This probability is known to the Sender, but not to the Receiver who updates his beliefs upon listening to Sender's announcements. If the Receiver participates in the project, he may learn the actual probability of success and has to decide whether to terminate the project getting a safe payoff or to keep working on it getting a random payoff. Following Koszegi and Rabin, 2006, 2007, 2009, we model the formation of reference points assuming that players have rational expectations about the game's play. In particular, we show that the Sender's equilibrium communication strategy can affect R's belief and can play a key-role in the formation of his reference point. This mechanism may lead R to keep working on projects that have low probability of success (in the previous example, R may insist to send the manuscript to a top general interest journal). We want to stress that this behavior is neither irrational, nor based

on a miscalculation of the actual probabilities of success. Indeed, once communication takes place and reference points are determined, R simply maximizes his total utility, that is the sum of consumption utility and gain/loss utility. Intuitively, if R initially trusted S and he subsequently finds out that she lied, he faces a trade-off: he can maximize his consumption utility by terminating the project and accepting the loss associated with a negative deviation from his reference utility or he can keep working on the project in the hope of decreasing (or even preventing) losses. If R is sufficiently loss averse, he will prefer the latter strategy even if he is fully aware that his action is suboptimal from the consumption utility point of view. Since this type of behavior may hurt S (in our example, R's stubbornness may lead to a rejection from a general-interest journal and delay the publication of the manuscript), she may prefer to tell the truth from the beginning.

The paper is organized as follows. The remainder of the Introduction reviews the relevant literature. In section 2 we present our model, we review reference dependence and loss aversion à la Köszegi and Rabin and we formalize the mechanism through which communication affects the formation of the reference point. Section 3 shows how reference dependence can lead to credible information transmission. To convey the main intuition of the paper and to highlight the linkage between reference point's formation and credible communication, we start imposing two simplifying assumptions: (i) some projects always succeed (in the previous example, some manuscripts are published on top general-interest journals with probability 1) and, (ii) upon initial participation, R always learns the project's probability of success. Under these assumptions, we show that S' announcement is credible and can affect R's behavior if and only if R is sufficiently loss averse and the behavior induced by a lie is sufficiently harmful for the Sender (in the previous example, if inefficient submissions are costly enough). Then, we relax the previous assumptions and show that our findings are robust, but new trade-offs arise. Section 3.2 shows that if all projects fail with some probability and loss aversion is high, R could prefer not to undertake the project even if it has high probability of success; this happens because R tries to avoid the loss associated with a failure (which, although unlikely, is still possible); as a result, S' announcement is credible and can affect R's behavior only if loss aversion is neither too high, nor too low. Section 3.3 further shows that in a model where R learns the true quality of the project only with some probability, truthtelling is possible either through the channel we described before or if the probability of finding out the project quality is sufficiently low. Section 4 discusses some specific features of our model and what would change if we modified them. Section 5 concludes.

1.1 Related Literature

This paper is mostly related to two strands of the literature: the one on strategic information transmission and the one on reference dependence; in what follows we provide an overview of the most relevant papers.

Starting from Crawford and Sobel, 1982, the literature on strategic communication has

looked for conditions under which information transmission is achievable.³ In line with this literature, we analyze a model in which the conflict of interests between S and R when the project has low probability of success prevents credible information transmission. However, we innovate on this literature assuming that R has reference-dependent preferences and we show how S' announcement may affect R's reference point and, consequently, S' optimal communication strategy.

Admittedly, credible communication can also be achieved through different channels. For instance, if S and R interact repeatedly, S' short term gain from lying could be lower than the cost induced by the loss in credibility that such a lie induces. In our model, S sends only one message and credible communication does not require long-lasting punishment strategies. Aumann and Hart, 2003 show that information transmission can be improved if we allow multiple rounds of communication, while Goltsman et al., 2009 and Ivanov, 2010 show that credible communication can be achieved introducing a mediator who weakens the link between S' announcements and R's response.⁴ In our model neither multiple rounds of communication, nor the introduction of a mediator would improve S' ability to transmit information. Finally, our model differs from the standard literature on cheap talk insofar we assume that R can detect S' lies during the game.⁵

The introduction of reference dependence and loss aversion in the economic literature dates back at least to Kahneman and Tversky, 1979 and it has been formalized by Kahneman and Tversky, 1991. Since then, reference dependence has been used in many papers leading to a debate concerning the actual formation of the reference point. Whereas some authors have taken a backward-looking approach assuming that the reference point is determined by an agent's status quo,⁶ other scholars have assumed that expectations about the future play a key role in determining the utility an agent feels entitled to.⁷ In this paper we follow the second approach and we show how communication may affect the formation of the reference point. In particular, using Köszegi and Rabin's, 2006, 2007, 2009 approach, we close the loop between reference point formation and optimal behavior through rational expectations: reference points are determined by agents' (correct) equilibrium beliefs and the optimality of a strategy is assessed taking into account its effect on the formation of reference point.⁸ Furthermore, our solution concept adapts those proposed by Köszegi and Rabin 2007, 2009

³Farrell and Rabin, 1996 and Krishna and Morgan, 2008 provide an overview of the literature. See also the references cited in footnote 1.

⁴The former paper deals with an unbiased mediator, while the latter allows for mediator's own bias.

⁵For more details on this assumption see Section 3.3. The assumption of verifiable announcements is analyzed in Dziuda, 2011, Ottaviani and Sorensen, 2006, Seidmann and Winter, 1997.

⁶See Kahneman et al., 1990, Kahneman and Tversky, 1991, Kahneman et al., 1991, Kahneman, 1992, Sugden, 2003 and the references therein.

⁷See Lleras, 2012, Ok et al., 2012, Sarver, 2012, Shalev, 2000 and the references cited below. Gul, 1991 provides a theory of disappointment aversion in which a lottery is evaluated based on how negative and positive outcome compare with its endogenously determined certainty equivalent.

⁸See Section 2.3 for further details.

to study the implications of S' equilibrium communication strategy on R's reference point.⁹ The relevance of opponents' behavior in the formation of reference points has been studied both theoretically and experimentally by Gill and Prowse, 2012 and Gill and Stone, 2010, who investigate tournament settings.

The link between communication and players' feelings, expectations and behavior has been recently analyzed by various authors. Koszegi, 2006 studies a principal-agent model in which the agent has some private information she would like to transmit to a principal who experiences anticipatory utility.¹⁰ Differently from our paper, Koszegi, 2006 allows for general anticipatory utilities, but focuses on the case in which the material interests of the two parties are perfectly aligned. By considering environments where players' interests are conflicting, we focus on the mechanism through which reference dependence and loss aversion can lead to credible information transmission.¹¹ Charness and Dufwenberg, 2006, 2010, 2011 provide theoretical and experimental evidence to show that communication may affect agent's participation in risky projects. Although our model share some features with theirs, the behavioral channel behind truth-telling is different. Whereas they look at the role of guilt aversion, we focus on reference dependence and loss aversion. This difference is not only semantic: our approach can be labelled as "independent of opponents' intentions", insofar R does not need to form beliefs on S' intentions; on the contrary, guilt aversion requires modelling players' higher-order belief about opponents' intentions.¹² Nevertheless, since we model a situation in which R's beliefs concerning his own future behavior affects his preferences over final outcomes, our paper belongs to the literature on psychological games, pioneered by Geanakoplos et al., 1989 and extended to dynamic environments by Battigalli and Dufwenberg, 2009.

Finally, some scholars have studied the implications of reference dependence and loss aversion in contractual settings. Fehr et al., 2011 and Hart and Moore, 2008 show that loss aversion can affect the choice between formal and informal contracting. In their model informal contracts guarantee higher flexibility, but their vagueness can induce an overestimation of future gains and, due to loss aversion, prevent agreements among the parties. de Meza and Webb, 2007 study the effect of loss aversion in a principal-agent model under different assumptions concerning the formation of the reference point. In Carbajal and Ely, 2012, a monopolist faces a continuum of loss averse consumers; in such a context, the authors show that the monopolist can exploit reference dependence and loss aversion to extract surplus from the consumers and they investigate how this possibility depends on the mechanism behind

⁹See also Koszegi, 2010. In particular, the "unexpected" events described in Koszegi and Rabin, 2007, 2009 correspond in our model to information sets in which R finds out the S lied when he was not expecting her to do so. Furthermore, in our model the probability of different states is affected by S' announcements.

¹⁰On anticipatory utilities see Loewenstein, 1987, Loewenstein and Prelec, 1992. For an axiomatic treatment see Caplin and Leahy, 2001, Epstein, 2008.

¹¹For a similar insight in a political economics setting, see Grillo, 2012

¹²On guilt aversion, see also Battigalli and Dufwenberg, 2007 and Battigalli et al., 2013. For the very same reason, our model differs from Rabin, 1994 interactive reciprocity, which could also lead to truth-telling.

reference point's formation.¹³ In section 4.3, we relate to this literature by discussing how enforceable monetary transfers could interact with loss aversion and help S establishing her credibility and affecting R's behavior.

2 The Model

2.1 The Game

Two agents, S (the Sender) and R (the Receiver), are involved in a project, whose probability of success can be either high, p_H , or low, p_L , with $p_L < p_H$. For simplicity, we refer to these projects as to "high-potential projects" and "standard projects". S knows the true quality of the project, but R does not and he assigns probability $\frac{1}{2}$ to each type of project.¹⁴ The timing is as follows: before the game starts, S learns the project's probability of success; we denote with H the state in which the project is high-potential and with L the state in which it is standard. Based on her information, S can send a message to R who has to decide whether to join the project (action *In*) or to stay out (action *Out*). In the former case R pays a participation fee equal to F and S gets an amount equal to G ; in the latter case the game is over and both agents get an outside payoff which we normalize to 0. If R plays *In*, he learns the true quality of the project and decides whether to invest an additional amount c in the project (action *Stay*) or to liquidate (action *Liquidate*) it ending the game without further changes in payoffs. If R plays *Stay*, the project succeeds (outcome s) or fails (outcome f) with probabilities determined by its potential; if the project succeeds, R gets 1 and S gets $W > 0$; if the project fails payoffs are respectively 0 and $L < 0$.¹⁵ Players do not discount the future and their initial utilities are normalized to 0. Figure 1 summarizes the structure of the game. We will denote with Z the set of terminal nodes of the game.

R's *consumption utility* depends both on the investment costs R incurs by participating in the project and on its outcome. Following the literature on reference dependence, we assume that R evaluates gains and losses in each of these two dimensions separately. This type of mental accounting could also be interpreted as capturing differences in the timing of payoffs: investments represent early payoffs, while outcome-related payoffs capture late payoffs.¹⁶ Although this mental accounting simplifies the analysis, Section 4.2 shows that under some additional condition on the representation of reference-dependent preferences the main insight of the paper would hold even with a unique dimension in R's utility.

¹³On the interaction between a monopolist and a mass of loss averse consumers, see also Rosato, 2012

¹⁴The insights of the paper would go through if we assumed that the *ex-ante* probability is $q \in (0, 1)$ and $1 - q$, but we would need to modify assumption 1 accordingly.

¹⁵Alternatively, we could assume S gets a positive return r_s (r_f) if the project succeeds (fails) and incurs a cost C if R chooses *Stay*. In this case, one could define $W = r_s - C$ and $L = r_f - C$ and interpret a failure as a situation in which the return on the project does not cover the cost C .

¹⁶Following this last interpretation, it would be reasonable to further distinguish between the initial fee F and the subsequent additional cost c . None of the results depends on the presence of this additional dimension in R's payoff.

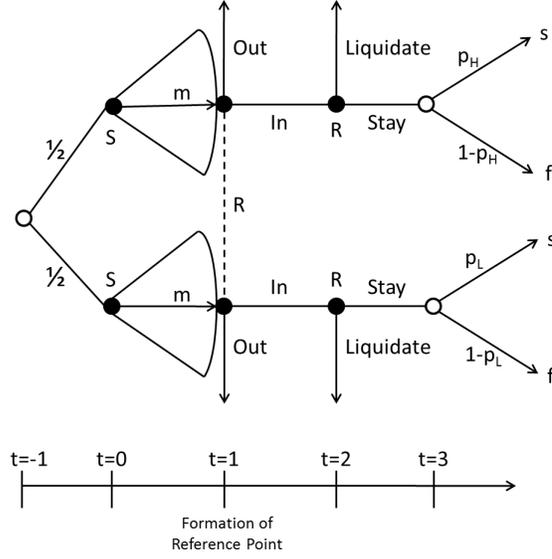


Figure 1: The Game

We make the following assumptions on payoffs.

Assumption 1 (i) $0 < p_L < c < p_H$, (ii) $F + c < p_H$, (iii) $\frac{p_H}{2} < F + \frac{c}{2}$.

Assumption 2 $G + p_H W + (1 - p_H) L > 0$.

Assumption 1(i) states that, conditional on having played *In*, R would liquidate standard projects and stay on high-potential ones.¹⁷ Assumptions 1(ii)-(iii) further imply that investment costs (F and c) are sufficiently low to guarantee R's participation if he is certain that the project is good (ii), but also sufficiently high to prevent his participation if he does not have additional information on the project's probability of success (iii).¹⁸ To put it differently, if S does not transmit any information, R will choose *Out*. Notice that playing *Stay* is both costly and risky: with some probability R gets a positive payoff, but with complementary probability, he gets nothing and the additional cost c is wasted. Finally, observe that $p_L > 0$: standard projects can still succeed with some probability. Assumption 2 simply states that S wants induce R's participation on high-quality projects.

Although stylized, the game represented in Figure 1 fits several environments.

¹⁷If either $p_L > c$ or $c > p_H$, R's choice between *Stay* and *Liquidate* would not depend on the project's probability of success and, as a result, S' information would not affect his participation in the project.

¹⁸If we assume that the *ex-ante* probability is q instead of $\frac{1}{2}$, this last assumption should become:

$$(iii') \quad qp_H < F + qc.$$

Example 1 *In the example we made in the introduction, F is the cost incurred by R in writing the manuscript (think of it as the opportunity cost of the time spent thinking about the problem, reading the relevant literature and writing the paper), while G is the net benefit S gets from a field journal publication (which is equal to 0 for R). Furthermore, c denotes the differential cost associated with a submission to a top general interest journal as opposed to a field journal (for instance, it can represent a more demanding revision). For simplicity, we can assume that this cost is constant for both researchers (this assumption can be easily generalized). Then, assumption 1(i) implies that players are willing to incur the differential cost c only if the project has high probability of being published in a top general interest journal.*

Example 2 *A child (R) has to decide whether to start playing the piano or not. The mother (S) may be better informed about the child's talent and his actual probability of becoming a good piano player (she may have talked with the music teacher or may be a music expert herself) and she may encourage him to start the activity, possibly overstating his actual probability of success. In this context, F represents an initial cost incurred by the child (e.g., giving up another activity or overcoming his laziness) and c is the cost associated with a continuous engagement in the activity (e.g., exercising several times per week, travelling to reach a better teacher). The mother is happy if the child learns some type of music (for instance, she may believe that music is an important component in her child education)); this benefit is captured by G . Furthermore, she experiences a positive (respectively, negative) utility of W (respectively, L) if the child becomes a good musician. W and L can be interpreted either as some type of empathic relationship with the child or, respectively, as the return if the child succeeds (W) and the cost incurred by supporting him (L).*

Example 3 *An entrepreneur (S) needs to raise money to finance a project. An investor (R) has the money, but is uncertain about the quality of the project, that is instead known to S . S may try to convince R that the project is good, but R has no way to ex-ante verify S ' statement. In this case, F represents the amount required to finance the project and c is an additional monetary disbursement needed to run the project (or alternatively the cost of effort exerted in the project). G represents the benefit that the entrepreneur gets from finding an investor (e.g., the amount of money S can divert to her own account if R finances the project or the gain she gets from relaxing the liquidity constraint). Similarly, W (L) represents S ' profit if the project succeeds (fails). In particular, a negative value of L can capture some sort of bankruptcy cost or simply a low return that does not cover the costs incurred by the entrepreneur.*

2.2 Strategies and Beliefs.

Let M be a finite set of messages. S ' behavior can be described with a function $s : \{L, H\} \rightarrow M$ that specifies a message for any possible information she may have. The set of mixed strategies is denoted with $\Sigma = \Delta(M^{\{L, H\}})$. Pure strategies will be denoted with s, s', s'' , while mixed strategies with $\sigma, \sigma', \sigma''$; we will abuse notation writing $s \in \Sigma$.

Since R is active both after S' initial announcement (when he has to choose between *In* or *Out*) and after having played *In* (in which case he learns the actual quality of the project and has to choose between *Stay* and *Liquidate*), his strategy can be represented by a behavioral strategy $m \mapsto \beta(m) = (e(m), x_L(m), x_H(m)) \in [0, 1]^3$, where $e(m)$ denotes the probability with which he chooses *In* if he receives message m and $x_Y(\cdot)$ represents the probability with which he chooses *Stay* if he receives message m , plays *In* and learns that the state is Y ($Y \in \{L, H\}$). Obviously, R's behavior depends on his beliefs concerning the project's probability of success. In particular, suppose R's conjecture about the communication strategy followed by S is represented by $\sigma \in \Sigma$; then if he receives message m , he will assign probability

$$\pi(m; \sigma) = \begin{cases} \frac{\sum_{s:s(H)=m} \sigma[s]}{\sum_{Y \in \{L, H\}} \sum_{s:s(Y)=m} \sigma[s]} & \text{if } \exists s : s(m) > 0 \\ \pi \in [0, 1] & \text{if } \forall s, s(m) = 0 \end{cases} \quad (1)$$

to the project being high potential.

Obviously, $\pi(m, H) = 1$ and $\pi(m, L) = 0$, where $\pi(m, Y)$ is the probability R assigns to the project being high potential, when he receives message m and finds that the state is Y (notice that these beliefs do not depend conjecture σ). We refer to $\{(\pi(m; \sigma), \pi(m, H), \pi(m, L))\}_{m \in M}$ as to the belief system associated with conjecture σ ; since $\pi(m, H), \pi(m, L)$ are trivially defined, we simply write $\{\pi(m; \sigma)\}_{m \in M}$.

2.3 Reference-Dependent Preferences

This paper uses Köszegi and Rabin's 2006, 2007, 2009 reference dependence model to capture the idea that agents care not only about final outcomes, but also about the comparison between these outcomes and an endogenously determined reference point. Let Z be a finite set of outcomes and consider a utility index $u : Z \rightarrow \mathbb{R}$. We say that an agent has reference-dependent utility if, for any pair of outcomes $a, r \in Z$, his utility is given by:

$$v(a | r) = u(a) + \mu(u(a) - u(r)), \quad (2)$$

where:

$$\mu(x) = \eta \cdot \max\{0, x\} + \eta\lambda \min\{0, x\} \quad \forall x \in \mathbb{R} \quad (3)$$

with $\eta \in (0, 1)$, and $\lambda > 1$. Thus, the utility of an agent is represented by a function $v(\cdot | \cdot) : Z \times Z \rightarrow \mathbb{R}$, where the first argument, a , is the actual outcome experienced by the agent and the second argument, r , is the reference outcome. In particular, utility index $u : Z \rightarrow \mathbb{R}$ represents the *consumption utility*, while $\mu : \mathbb{R} \rightarrow \mathbb{R}$ captures the *gain/loss utility*: whenever the utility associated with outcome a , $u(a)$, exceeds (respectively, falls short of) the reference

utility $u(r)$, the agent experiences a gain (respectively, loss). In this setting η measures the relative weight of the gain/loss utility compared to the consumption utility, while $\lambda > 1$ captures *loss aversion*, namely the fact that R dislikes losses more than he likes equal-size gains.

The utility function can be extended to random outcomes and random reference points as follows: let $\tilde{a} \in \Delta(Z)$ and $\tilde{r} \in \Delta(Z)$, then:

$$v(\tilde{a} | \tilde{r}) = \sum_{r \in Z} \sum_{a \in Z} v(a | r) \tilde{a}[a] \tilde{r}[r]$$

Finally, suppose that the space of outcomes is multidimensional: $Z = Z_1 \times Z_2 \times \dots \times Z_n$; $z \in Z$ is a n -dimensional vector $(z_1, z_2, \dots, z_n) \in Z$. For each pair $\tilde{a}, \tilde{r} \in \Delta(Z)$, total utility is given by:

$$v(\tilde{a} | \tilde{r}) = \sum_{r \in Z} \sum_{a \in Z} \left(\sum_{i=1}^n v_i(a_i | r_i) \right) \tilde{a}[a] \tilde{r}[r], \quad (4)$$

where for every $i = 1, 2, \dots, n$, $v_i(a_i | r_i)$ is defined as in (2). Note that (4) implies that R has separable and additive utilities in the n dimensions.

Following, Köszegi and Rabin 2006, 2007, 2009 we endogenize the formation of the reference point through equilibrium analysis. To understand this point, consider a static decision problem in which a decision maker has to choose an element from a finite set D . Let $\zeta : D \rightarrow Z$ be the *outcome function*. Thus, whenever the decision maker chooses d , she foresees that the outcome will be $\zeta(d)$ and, consequently, $\zeta(d)$ becomes her reference point. In this context, two definitions of optimality are possible: the first evaluates deviations without taking into account their effect on the reference point, while the latter accounts for the change in the reference point induced by the deviation.¹⁹

In the model we are analyzing the first type of optimality can be interpreted as a consistency requirement, while the second one is a criterion that selects the strategy with the highest utility. To be more precise, suppose R's conjecture concerning S' communication strategy is σ and consider strategy $\beta(\cdot) = (e(\cdot), x_L(\cdot), x_H(\cdot))$. Then for every message m , $\pi(m, \sigma)$ and $\beta(m)$ induce a distribution over final outcomes; this distribution represents R's beliefs about the evolution of the game conditional on him following strategy β . We denote it with $\tilde{\zeta}_M(\beta; m, \sigma)$ and we assume it constitutes his reference point. Thus, the reference point can be regarded as a function of m ($m \mapsto \tilde{\zeta}_M(\beta; m, \sigma)$) and represented in the following tables:

Terminal node: z	Probability z is reached: $\tilde{\zeta}_M(\beta; \cdot, \sigma)[z]$
$\{L, Out\}$	$(1 - \pi(\cdot, \sigma))(1 - e(\cdot))$
$\{L, In, Liquidate\}$	$(1 - \pi(\cdot, \sigma))e(\cdot)(1 - x_L(\cdot))$
$\{L, In, Stay, Success\}$	$(1 - \pi(\cdot, \sigma))e(\cdot)x_L(\cdot)p_L$
$\{L, In, Stay, Failure\}$	$(1 - \pi(\cdot, \sigma))e(\cdot)x_L(\cdot)(1 - p_L)$

¹⁹Conditions guaranteeing that optimality is well defined are provided in Köszegi, 2010.

Terminal node: z	Probability z is reached: $\tilde{\zeta}_M(\beta; \cdot, \sigma)[z]$
$\{H, Out\}$	$\pi(\cdot, \sigma)(1 - e(\cdot))$
$\{H, In, Liquidate\}$	$\pi(\cdot, \sigma)e(\cdot)(1 - x_H(\cdot))$
$\{H, In, Stay, Success\}$	$\pi(\cdot, \sigma)e(\cdot)x_H(\cdot)p_H$
$\{H, In, Stay, Failure\}$	$\pi(\cdot, \sigma)e(\cdot)x_H(\cdot)(1 - p_H)$

(we do not specify the message sent by S in the description of final outcomes because it does not affect R's consumption utility). Obviously each $\tilde{\zeta}_M(\beta; m, \sigma)$ also induces a distribution over consumption utilities in each dimension:

$$\tilde{u}\left(\tilde{\zeta}_M(\beta; \cdot, \sigma)\right) = \left(\tilde{u}_1\left(\tilde{\zeta}_M(\beta; \cdot, \sigma)\right), \tilde{u}_2\left(\tilde{\zeta}_M(\beta; \cdot, \sigma)\right)\right),$$

where:

$$\tilde{u}_1\left(\tilde{\zeta}_M(\beta; \cdot, \sigma)\right)[w] = \begin{cases} e(\cdot)(\pi(\cdot, \tilde{s})x_H(\cdot)p_H + (1 - \pi(\cdot; \sigma))x_L(\cdot)p_L) & \text{if } w = 1 \\ 1 - e(\cdot)(\pi(\cdot; \sigma)x_H(\cdot)p_H + (1 - \pi(\cdot; \sigma))x_L(\cdot)p_L) & \text{if } w = 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\tilde{u}_2\left(\tilde{\zeta}_M(\beta; \cdot, \sigma)\right)[w] = \begin{cases} 1 - e(\cdot) & \text{if } w = 0 \\ e(\cdot)(\pi(\cdot; \sigma)(1 - x_H(\cdot)) + (1 - \pi(\cdot; \sigma))(1 - x_L(\cdot))) & \text{if } w = -F \\ \pi(\cdot; \sigma)e(\cdot)x_H(\cdot) + (1 - \pi(\cdot; \sigma))e(\cdot)x_L(\cdot) & \text{if } w = -F - c \\ 0 & \text{otherwise} \end{cases}.$$

It is important to point out that $x_L(\cdot), x_H(\cdot)$ are not always relevant in determining the reference point; for instance, if $e(m) = 0$ R's reference point depends neither on $x_L(m)$ nor on $x_H(m)$; similarly, if $\pi(m; \sigma) = 1$ (respectively, $\pi(m; \sigma) = 0$), the reference point depends only on $x_H(m)$ (respectively, $x_L(m)$).

Strategy β also induces a distribution over final outcomes after that R finds out that the state is $Y \in \{L, H\}$.²⁰ For any $Y \in \{L, H\}$, this distribution is denoted with $\tilde{\zeta}_Y(\beta; Y)$ and

²⁰Given our assumption about belief updating, this distribution does not depend on the message sent by S.

can be represented as follows:

Terminal node: z	Prob. z is reached: $\tilde{\zeta}_Y(\beta; Y)[z]$
$\{Y, In, Stay, Success\}$	$x_Y(\cdot) p_Y$
$\{Y, In, Stay, Failure\}$	$x_Y(\cdot) (1 - p_Y)$
$\{Y, In, Liquidate\}$	$1 - x_Y(\cdot)$
other terminal nodes	0

Obviously such a distribution depends neither on $e(\cdot)$ nor on $x_{Y'}(\cdot)$ with $Y' \neq Y$. It is immediate to see that $\tilde{\zeta}_Y(\beta; y)$ induces a distribution over final utilities given by:

$$\tilde{u}(\tilde{\zeta}_Y(\beta; Y)) = \left(\tilde{u}_1(\tilde{\zeta}_Y(\beta; Y)), \tilde{u}_2(\tilde{\zeta}_Y(\beta; Y)) \right), Y \in \{L, H\}$$

where

$$\tilde{u}_1(\tilde{\zeta}_Y(\beta; y))[w] = \begin{cases} x_Y(m) p_y & \text{if } w = 1 \\ 1 - x_Y(m) p_y & \text{if } w = 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\tilde{u}_2(\tilde{\zeta}_Y(\beta; y))[w] = \begin{cases} x_Y(m) & \text{if } w = -F - c \\ 1 - x_Y(m) & \text{if } w = -F \\ 0 & \text{otherwise} \end{cases}$$

For any two finite lotteries over real number, \tilde{u} and \tilde{v} , let:

$$\mathbb{E}\tilde{u} = \sum_{x \in \mathbb{R}} \tilde{u}[x] \cdot x$$

and:

$$\mu(\tilde{u} - \tilde{v}) = \sum_{x \in \mathbb{R}} \sum_{y \in \mathbb{R}} \mu(\tilde{u}[y]y - \tilde{v}[x]x).$$

We are now ready to define R's *total utility* function. If R receives message m , decides to follow strategy β , his reference point is based on strategy β' and he believes S is following communication strategy σ , his total utility will be given by:

$$v\left(\tilde{\zeta}_M(\beta; m, \sigma) \mid \tilde{\zeta}_M(\beta'; m, \sigma)\right) = \mathbb{E}\tilde{u}\left(\tilde{\zeta}_M(\beta; m, \sigma)\right) + \mu\left(\tilde{u}\left(\tilde{\zeta}_M(\beta; m, \sigma)\right) - \tilde{u}\left(\tilde{\zeta}_M(\beta'; m, \sigma)\right)\right);$$

Similarly, R's total utility when he finds out that the state is $Y \in \{L, H\}$, decides to follow strategy β and his reference point is based on strategy β' , message m and conjecture σ is given

by:

$$v\left(\tilde{\zeta}_Y(\beta; Y) \mid \tilde{\zeta}_M(\beta'; m, \sigma)\right) = \mathbb{E}\tilde{u}\left(\tilde{\zeta}_Y(\beta; Y)\right) + \mu\left(\tilde{u}\left(\tilde{\zeta}_Y(\beta; Y)\right) - \tilde{u}\left(\tilde{\zeta}_M(\beta'; m, \sigma)\right)\right).$$

Notice that in this latter case the reference point is still the one established after announcement $m \in M$. This assumption captures the idea that S' announcement has some lasting effect on the formation of the reference point. See section 4.5 for further details on this.

Now suppose R's reference point is determined based on strategy β' . In equilibrium, we impose the following consistency requirement: if the reference point is given by $\tilde{\zeta}_M(\beta'; m, \sigma)$, R must be willing to follow the behavior prescribed by β' at every information set he may encounter. Strategies that satisfy this requirement are called *reference-point-consistent strategies*.

Definition 1 Fix a set of messages M . $\beta'(\cdot) = (e'(\cdot), x'_L(\cdot), x'_H(\cdot))$ is a reference-point-consistent strategy given σ at message $m \in M$ if for every $\beta(\cdot)$:

- (i) $v\left(\tilde{\zeta}_M(\beta'; m, \sigma) \mid \tilde{\zeta}_M(\beta'; m, \sigma)\right) \geq v\left(\tilde{\zeta}_M(\beta; m, \sigma) \mid \tilde{\zeta}_M(\beta'; m, \sigma)\right)$;
- (ii) $v\left(\tilde{\zeta}_Y(\beta'; Y) \mid \tilde{\zeta}_M(\beta'; m, \sigma)\right) \geq v\left(\tilde{\zeta}_Y(\beta; Y) \mid \tilde{\zeta}_Y(\beta'; m, \sigma)\right) \quad \forall Y \in \{L, H\}$.

A strategy β is reference-point-consistent given σ if it is reference-point-consistent given σ at every message $m \in M$.

Thus, reference-point consistency evaluates deviations without taking into account the effect of these deviations on the reference point. However, Definition 1 does not rule out the existence of multiple reference-point-consistent strategies associated with different total utilities. The following definition selects among them:²¹

Definition 2 Fix a set of messages M . β^* is a preferred personal strategy (PPS) given σ if:

- (i) β^* is a reference-point-consistent given σ and, (ii) for any other reference-point-consistent strategy given σ , β , we have that for every message $m \in M$:

$$v\left(\tilde{\zeta}_M(\beta^*; m, \sigma) \mid \tilde{\zeta}_M(\beta^*; m, \sigma)\right) \geq v\left(\tilde{\zeta}_M(\beta; m, \sigma) \mid \tilde{\zeta}_M(\beta; m, \sigma)\right). \quad (5)$$

Thus, the definition of a preferred personal strategy implies R will select one of the dynamic consistent strategies that maximize *ex-ante* utility after every message m . This definition deserves some additional comments. First of all, given the definition of reference-point consistency, a sufficient condition for the existence of a preferred personal strategy is the existence of finitely many reference-point-consistent strategies for every message m ; in our model, this is indeed the case. Also notice that the definition of PPS evaluates the optimality of a strategy only at the *ex-ante* stage; however, this should not be interpreted as a failure of some generalized version of subgame perfection: condition (ii) in the definition of

²¹On preferred personal equilibria, see Koszegi, 2010 and Koszegi and Rabin, 2006, 2007, 2009.

reference-point consistency already imposes this type of perfection. Instead, adding a similar requirement in the definition of PPS would rule out strategies based on a comparison between continuation utilities that never actually arises. Finally, although the definition of PPS seems to suggest that the choice of the reference point is a conscious act, we could also interpret it as the result of some unconscious process.²² Furthermore, the assumption that R selects the optimal strategy can be relaxed; for instance, we could assume R randomizes among all reference-point-consistent strategies and none of the qualitative results would be affected.

We assume that the Sender has standard expected utility based on utility indexes $g(z)$ associated with terminal nodes. A discussion of what would happen if S had reference-dependent preferences is provided in Section 4.4. Let $\tilde{\zeta}_S(m, \beta; Y)$ denote the distribution over final outcomes induced by message m , when R follows strategy β and the state is Y . Then S' expected utility when the state is Y , she sends message m and R follows strategy β can be written as:

$$g(m, \beta; Y) = \sum_{z \in Z} \tilde{\zeta}_S(m, \beta; Y)[z] \cdot g(z)$$

We will analyze the model using the following equilibrium definition:

Definition 3 *Fix a set of message M . A triple $(\sigma^*, \beta^*, \pi^*) \in \Sigma \times [0, 1]^3 \times [0, 1]^M$ is an equilibrium if:*

- (i) β^* is R's preferred personal strategy given σ^* ;
- (ii) if $\sigma^*[s] > 0$, then for every state $Y \in \{L, H\}$ and every message $m \in M$, $g(s(Y), \beta^*; Y) \geq g(m, \beta^*; Y)$;
- (iii) for every m , $\pi^*(m) = \pi(m; \sigma^*)$.

This paper focuses on two classes of equilibria: those in which S' announcement does not affect R's behavior and those in which S' fully reveals her information and her message modifies R's behavior. The formal definition of these classes is as follows:

Definition 4 *Fix a set of messages M . Then:*

- $(\sigma^*, \beta^*, \pi^*)$ is a message-independent equilibrium if for every $m \in M$, $e^*(m) = 0$.
- $(\sigma^*, \beta^*, \pi^*)$ is a fully informative and message-dependent equilibrium if it is an equilibrium, there exists a message m such that $e^*(m) > 0$ and

$$\{m : \exists s, \sigma^*(s) > 0, s(H) = m\} \cap \{m : \exists s, \sigma^*(s) > 0, s(L) = m\} = \emptyset;$$

Section 4.1 shows that the assumption of full information transmission is without loss of generality and the only relevant distinction is between message-dependent and message-independent equilibria. We conclude this section pointing out that if R does not have reference-dependent preferences ($\eta = 0$), our equilibrium definition coincides with standard ones.

²²On this point, see the discussion in Brunnermeier and Parker, 2005

Remark 1 If $\eta = 0$, $(\sigma^*, \beta^*, \pi^*)$ is an equilibrium if and only if it is a sequential equilibrium.

3 Reference Dependence and Credible Communication

3.1 Risk-free High-Potential Projects

In standard models of communication, S' message may affect R's behavior and his consumption utility, $\mathbb{E}\tilde{u}\left(\tilde{\zeta}_M(\beta; m, \sigma)\right)$, by changing the probability weight that R assigns to high-potential projects, $\pi(m; \sigma)$, and, consequently, to the terminal nodes. Although in our model this channel is still active, a new one arises: message m may affect the formation of the reference point and, as a result, the gain/loss utility $\mu\left(\tilde{u}\left(\tilde{\zeta}_M(\beta; m, \sigma)\right) - \tilde{u}\left(\tilde{\zeta}_M(\beta; m, \sigma)\right)\right)$. This additional channel plays a key role in our analysis and will enable S to transmit her information truthfully. To highlight the mechanism through which S' credibility can be attained, we make the following assumption that we will relax in Section 3.2.

Assumption 3 High quality projects always succeed: $p_H = 1$.

We begin our analysis characterizing the equilibria under the assumption that R does not have reference-dependent preferences ($\eta = 0$).

Proposition 1 Let $\eta = 0$. If assumptions 1, 2 and 3 hold, then all equilibria are message-independent and R's preferred personal strategy is such that for every message $m \in M$, $(e(m), x_L(m), x_H(m)) = (0, 0, 1)$.

Proof. Assumption 1(i) implies that in any equilibrium, conditional on having played In, R would choose Liquidate if the project has low probability of success and Stay if such probability is high. Formally, for every $m \in M$, $x_L(m) = 0$ and $x_H(m) = 1$. Thus, if R plays In after message m , S gets $W + G > 0$ if the state is H and $W > 0$ if the state is L. Consider communication strategy $s(L) = s(H) = m'$ and, for any message $m \in M \setminus \{m'\}$, set $\pi(m; s) = \frac{1}{2}$ and $\beta(m) = (0, 0, 1)$. It is easy to verify that this is an equilibrium. Pick any other equilibrium and denote with σ^* and β^* the players' strategies in such an equilibrium. By the previous argument for every $m \in M$, $x_L^*(m) = 0$ and $x_H^*(m) = 1$ Let

$$M^\circ(\sigma^*) = \{m \in M : \exists s \in \Sigma, \exists Y \in \{L, H\}, \sigma^*[s] > 0, s(Y) = m\};$$

$M^\circ(\sigma^*)$ is the set of messages that arise with positive probability when S follows strategy σ^* . Take $\bar{m} \in \arg \max_{m \in M^\circ(\sigma^*)} e^*(m)$. If $e^*(\bar{m}) > 0$, then for every $m \in M$ and $Y \in \{L, H\}$, $g(m, \beta^*; Y) > 0$; indeed S can always send \bar{m} regardless of her actual information and get $e^*(\bar{m})(G + \frac{W}{2})$ in state H and $e^*(\bar{m})G$ in state L. Furthermore, $e^*(\bar{m}) > 0$ and assumption 1(iii) implies $\pi(\bar{m}; \sigma^*) > \frac{1}{2}$. Thus, $\sum_{s:s(H)=\bar{m}} \sigma^*[s] > \sum_{s:s(L)=\bar{m}} \sigma^*[s]$. Then, we can find another message $\underline{m} \in M^\circ(\sigma^*)$ such that $\sum_{s:s(H)=\underline{m}} \sigma^*[s] < \sum_{s:s(L)=\underline{m}} \sigma^*[s]$, so that

$\pi(\underline{m}; \sigma^*) < \frac{1}{2}$. Assumption 1(iii) implies $e(\underline{m}) = 0$, and consequently $g(\underline{m}, \beta^*; Y) = 0$. This establishes a contradiction. We conclude that in equilibrium $e(m) = 0$ for every $m \in M$. ■

Thus, without reference dependence S cannot credibly state that the project has high probability of success and this prevents R's participation. Such an inefficiency is due to a two-sided lack of commitment power: on the one hand, S cannot commit to reveal her private information truthfully; on the other hand, R cannot commit to punish S if she deviates from truthtelling. To put it differently, since S does not incur any cost from lying (either directly because he values honesty or indirectly through an endogenous punishment carried out by R), she will always have an incentive to claim that the project is a high-potential one. This, in turn, destroys her credibility and leads R to choose *Out* independently of the message received.

Now, suppose that R has reference-dependent preferences and is loss averse; then, his utility is represented by function $v(\cdot | \cdot)$. In the remaining of this section we will show that a message-dependent equilibrium may arise. In such an equilibrium S truthfully reveals her private information and R believes her behaving accordingly. Assume, without loss of generality, that the set of messages is given by $M = \{L, H\}$;²³ message $m = L$ (H) is interpreted as "the project has low (high) probability of success". Denote the fully informative strategy with s^T ; thus, $s^T(L) = L$ and $s^T(H) = H$.

To prove the existence of a fully informative equilibrium, we proceed in two steps: first, we compute R's preferred personal strategy given $s^T, \beta^T(\cdot)$. Then, we verify that s^T is indeed optimal given $\beta^T(\cdot)$.

Proposition 2 *Let $\eta > 0$. If assumptions 1 and 3 hold, then R's preferred personal strategy given s^T is $e^T(L) = 0, e^T(H) = 1$,*

$$x_L^T(m) = \begin{cases} 0 & \text{if } m = L \\ \chi_L(\lambda) & \text{if } m = H \end{cases}, \quad x_H^T(m) = \begin{cases} \chi_H(\lambda) & \text{if } m = L \\ 1 & \text{if } m = H \end{cases}$$

where

$$\chi_L(\lambda) = \begin{cases} 0 & \text{if } \lambda < \frac{c(1+\eta)-pL}{\eta pL} \\ x \in [0, 1] & \text{if } \lambda = \frac{c(1+\eta)-pL}{\eta pL} \\ 1 & \text{if } \lambda > \frac{c(1+\eta)-pL}{\eta pL} \end{cases}, \quad \chi_H(\lambda) = \begin{cases} 0 & \text{if } \lambda > \frac{1+\eta-c}{\eta c} \\ x \in [0, 1] & \text{if } \lambda = \frac{1+\eta-c}{\eta c} \\ 1 & \text{if } \lambda < \frac{1+\eta-c}{\eta c} \end{cases}$$

Proof. Throughout this proof, we omit to specify the dependence of variables on s^T . Let $m = H$. Since $\pi(H) = 1$, $x_L(H)$ does not affect the formation of the reference point.

²³In particular, we can assume that R interprets any other message $m \notin \{L, H\}$ as if S sent message L .

Denote with β^* a strategy whose continuation strategy after message $m = H$ is given by $(1, \chi_L(\lambda), 1)$. If R were planning to follow this strategy, his reference utility after message $m = H$ would be given by:

$$\tilde{u}_1\left(\tilde{\zeta}_M(\beta^*; H)\right)[w] = \begin{cases} 1 & \text{if } w = -F - c \\ 0 & \text{otherwise} \end{cases}, \quad \tilde{u}_2\left(\tilde{\zeta}_M(\beta^*; H)\right)[w] = \begin{cases} 1 & \text{if } w = 1 \\ 0 & \text{otherwise} \end{cases}$$

It is easy to verify that $v\left(\tilde{\zeta}_M(\beta^*; H) \mid \tilde{\zeta}_M(\beta^*; H)\right) = 1 - F - c$. Notice that keeping the reference utility fixed, any strategy that prescribes $e(H) = q < 1$ would lead at most to a total utility equal to $q(1 - F - c) - (1 - q)\eta\lambda + (1 - q)(F + c)$ which is lower than $\left(\tilde{\zeta}_M(\beta^*; H) \mid \tilde{\zeta}_M(\beta^*; H)\right)$ by assumption 1(ii). Furthermore, under this reference utility, $x_L(H) = \chi_L(\lambda)$ and $x_H(H) = 1$ maximizes total continuation utility after In .²⁴ We conclude that β^* is reference-point-consistent at $m = H$ given s^T . We now show that preferred personal strategies must prescribe $\beta^*(H)$ after message H . To see why, notice that a strategy that prescribes $e(H) = 0$ would lead a utility equal to $0 < 1 - F - c$. Therefore a preferred personal strategy must prescribe $e(H) > 0$. Suppose $e(H) > 0$ and $x_H(H) < 1$. In this case, R 's utility would be equal to:

$$-F + \eta ce(H)x_H(H) - \eta\lambda e(H)x_H(H) - \eta\lambda c(1 - e(H))$$

which is negative by assumption 1(i). Thus, whenever a preferred personal strategy prescribes participation with positive probability, it must also prescribe $x_H(H) = 1$. Thus, the only possibility $e(H) \in (0, 1)$ and $x_H(H) = 1$. Under this strategy, R 's total utility would be equal to $0 - e(H)(\eta\lambda - \eta(F + c))$, which is negative by assumption 1(ii). We conclude that a preferred personal strategy must prescribe continuation strategy $(1, \chi_L(\lambda), 1)$ after message $m = H$.

Now, let $m = L$. Since $\pi(L) = 0$, $x_H(L)$ does not affect the formation of the reference point. Let β^* be a strategy whose continuation strategy after message $m = L$ is given by $\beta^*(L) = (0, 0, \chi_H(\lambda))$. If R were planning to follow this strategy, his reference utility after message $m = L$ would be given by:

$$\tilde{u}_1\left(\tilde{\zeta}_M(\beta^*; L)\right)[w] = \begin{cases} 1 & \text{if } w = 0 \\ 0 & \text{otherwise} \end{cases}, \quad \tilde{u}_2\left(\tilde{\zeta}_M(\beta^*; L)\right)[w] = \begin{cases} 1 & \text{if } w = 0 \\ 0 & \text{otherwise} \end{cases}$$

Furthermore, $v\left(\tilde{\zeta}_M(\beta^*; L) \mid \tilde{\zeta}_M(\beta^*; L)\right) = 0$. Given this reference utility any strategy that

²⁴ $x_H(H) = 1$ follows from assumption 1(i). $x_L(H) = \chi_L(\lambda)$ follows since under this reference utility, R 's continuation utility from playing *Stay* is given by $1 - F - c - \eta\lambda(1 - p_L)$, while his utility from playing *Liquidate* is given by $-F + \eta c - \eta\lambda$.

continuation strategy $(q, x_L(L), x_H(L))$ after message $m = L$ would induce a total utility equal to:

$$qx_L(L)p_L - qF - qx_L(L)c + \eta qx_L(L)p_L - \eta \lambda qF - \eta \lambda qx_L(L)c. \quad (6)$$

Assumption 1(i) implies that (6) is negative for every $q \in (0, 1]$. In addition, under this reference utility, one can verify that $x_L(L) = 0$ and $x_H(L) = \chi_H(\lambda)$ maximize continuation utility after In.²⁵ Thus β^* is reference-point consistent at $m = L$ given s^T . Now, consider any strategy prescribing $(\hat{e}(L), \hat{x}_L(L), \hat{x}_H(L))$ after message $m = L$. If $\hat{e}(L) = 0$, we have already characterized the unique reference-point consistent strategy. Let $\hat{e}(L) = 1$. In this case, expected total utility would be given by:

$$\hat{x}_L(L)p_L - F - \hat{x}_L(L)c - \eta(\lambda - 1)\hat{x}_L(L)(1 - \hat{x}_L(L))c - \eta(\lambda - 1)p_L\hat{x}_L(L)(1 - p_L\hat{x}_L(L))$$

which is lower than 0 by assumption 1(i). Thus, even if this strategy were reference-point-consistent, it would not be a preferred personal strategy. Suppose, instead, $\hat{e}(L) \in (0, 1)$. In this case, R must be indifferent between playing In or playing Out after message $m = L$. Thus, the following equality must hold:

$$\begin{aligned} \hat{x}_L(L)(p_L - c + \eta\lambda\hat{e}(L)\hat{x}_L(L)p_L^2 + \eta p_L(1 - \hat{e}(L)\hat{x}_L(L)p_L) - \eta\lambda(1 - \hat{e}(L)\hat{x}_L(L))c) = \\ = F + \eta\lambda(1 - \hat{e}(L))F + \eta\hat{x}_L(L)\hat{e}(L)\hat{x}_L(L)c + \eta\hat{e}(L)F \end{aligned}$$

If $\hat{x}_L(L) = 0$, the previous condition becomes $-F - \eta\lambda(1 - \hat{e}(L))F = \eta\hat{e}(L)F$, which cannot be satisfied. If $\hat{x}_L(L) \in (0, 1)$, reference-point consistency would further require:

$$p_L - c + \eta\lambda\hat{e}(L)\hat{x}_L(L)p_L^2 + \eta p_L(1 - \hat{e}(L)\hat{x}_L(L)p_L) - \eta\lambda(1 - \hat{e}(L)\hat{x}_L(L))c = \eta\hat{e}(L)\hat{x}_L(L)c.$$

Thus, the indifference condition between In and Out would imply:

$$F + \eta\lambda(1 - \hat{e}(L))F + \eta\hat{e}(L)F = 0$$

which cannot be satisfied. Finally, if $\hat{x}_L(L) = 1$ R 's total utility from entering would be equal to:

$$p_L - F - c + \eta p_L(1 - \hat{e}(L)p_L) - \eta\lambda(1 - p_L)\hat{e}(L)p_L - \eta\lambda(1 - \hat{e}(L))(F + c) < 0$$

where the inequality follows from assumption 1(ii) and $\lambda > 1$. Thus a preferred personal strategy cannot prescribe $\hat{e}(L) \in (0, 1)$ and $\hat{x}_L(L) = 1$ after message $m = L$.

We conclude that the (unique) preferred personal strategy β^* is the one described in the state-

²⁵ $x_L(L) = 0$ follows immediately from assumption 1(i). $x_H(L) = \chi_H(\lambda)$ comes from comparing R 's total utility if he play *Stay* after finding out that the project has high probability of success, $1 - F - c - \eta\lambda F + \eta - \eta\lambda c$, with the total utility associated with *Liquidate*, $-F - \eta\lambda F$.

ment of the proposition. ■

An immediate corollary of proposition 2 is that the reference point associated with the unique preferred personal strategy given s^T would be a certain success after message $m = H$ and no participation after message $m = L$. Furthermore, the dependence of the reference point on m , implies that R's choice between *Stay* and *Liquidate* depends on (i) the coefficient of loss aversion, and (ii) the message sent by S. This feature is particularly relevant after message $m = H$; the next corollary focuses on this case.

Remark 2 Under the assumption of proposition 2, there exists a threshold level $\lambda^*(p_L, c, \eta) = \frac{c(1+\eta)}{p_L\eta} - \frac{1}{\eta}$ such that R's preferred personal strategy given s^T prescribes $e(H) = 1$ and $x_i(H) = 1$ for every $i \in \{L, H\}$ if and only if $\lambda \geq \lambda^*(p_L, c, \eta)$.

Thus if the coefficient of loss aversion is sufficiently high, R's preferred personal strategy at $m = H$ given s^T would require him to play *In* and to keep working on the project even if he finds out that it has low probability of success. This happens even though, as far as consumption utility is concerned, R would be better off liquidating the project. The intuition behind this result is as follows. In a truthtelling equilibrium, if S sends message $m = H$, R gets acquainted with the idea of getting a large final payoff incurring all the relevant costs. Therefore, if he liquidates the project, he incurs a loss associated with a payoff lower than expected, while if he plays *Stay*, he can still attain his reference utility (this happens with probability p_L or p_H). To put it differently, if R finds out that S claimed that the project is high-potential even though it is not, he would face a trade-off: he could take the action that maximizes his consumption utility but yields losses (*Liquidate*), or he could try to reduce his loss, by playing an action that is suboptimal from the consumption utility point of view (*Stay*). Obviously, R's actual behavior depends on the relative importance of loss aversion: as remark 2 points out, *Stay* will be chosen whenever λ is sufficiently high.

Notice that R's decision to play *Stay* even when the probability of success is p_L is due to a change in R's preferences induced by S' message and by R's own optimizing behavior. Indeed, R's choice between *Liquidate* and *Stay* can be described as the choice between two lotteries on \mathbb{R}^2 . *Liquidate* is equivalent to choosing a lottery that delivers $(-F - c, 0)$ with probability 1, while *Stay* is equivalent to a lottery that delivers $(-F - c, 1)$ with probability p_Y and $(-F - c, 0)$ with probability $(1 - p_Y)$, $Y \in \{L, H\}$. Since $p_L < c$, an agent immune from reference dependence who has separable and additive preferences over the two dimensions, would choose *Liquidate*. However, if we introduce reference dependence, S' message could induce reference utility $(-F - c, 1)$; if this happens, negative deviations in the second dimension would be evaluated with weight $\eta\lambda$, while positive deviation in the first dimension would have weight $\eta < \eta\lambda$. Therefore, as we increase λ , *Stay* becomes more and more appealing and will eventually be chosen over *Stay*; this will happen exactly when λ exceed $\lambda^*(p_L, c, \eta)$.

Intuitively, the worse in terms of consumption utility action *Stay* is, the higher the threshold level for loss aversion, $\lambda^*(p_L, c, \eta)$, will be. The following remark makes this statement precise.

Remark 3 $\lambda^*(p_L, c, \eta)$ is increasing in c and decreasing in p_L and η .

Summarizing, if R believes that S is being sincere, loss aversion may induce him to react to a lie by choosing *Stay* even if he knows that the project has low probability of success. This may have important implications for S' optimal communication strategy.

Proposition 3 Let $\eta > 0$. If assumptions 1, 2 and 3 hold, then a a fully informative and message-dependent equilibrium exists if and only if $\lambda \geq \lambda^*(p, c, \eta)$ and $L < -\frac{G+p_L W}{1-p_L}$.

Proof. Suppose $\lambda \geq \lambda^*(p, c, \eta)$, $L < -\frac{G+p_L W}{1-p_L}$. We will construct a fully-revealing and message-dependent equilibrium. If R thinks S is following strategy s^T , R's behavior is described by Proposition 2. Then, if the project is high-potential, S would get a utility equal to $W + G$ by sending message $m = H$ and equal to 0 by sending message $m = L$. Obviously, she will send the former message. If, instead, the project is standard, S would get a utility equal to 0 by sending message $m = L$ and equal to $p_L G + (1 - p_L) L + W$ if $m = H$ (in this last case, we are breaking indifference assuming that $\chi_L(\lambda^*(p, c, \eta)) = 1$). Since $L < -\frac{G+p_L W}{1-p_L}$, S will prefer to send message $m = L$. Thus a fully-revealing and message-dependent equilibrium exists.

Suppose that a fully-revealing and message-dependent equilibrium exists. Then S is following communication strategy s^T and proposition 2 characterizes R's preferred personal strategy given s^T . Since this equilibrium is fully revealing, S must reveal her private information in state L, and this requires $x_L(H)(p_L W + (1 - p_L) L) + G \leq 0$. The previous inequality can be satisfied only if $L < -\frac{G+p_L W}{1-p_L}$ and $\lambda \geq \lambda^*(p, c, \eta)$ (recall that $G > 0$ and that proposition 2, implies that R's preferred personal strategy given s^T prescribes $x_L(H) > 0$ only if $\lambda \geq \lambda^*(p, c, \eta)$). ■

Thus, thanks to reference dependence and loss aversion a fully-revealing and message-dependent equilibrium may exist even though this is not the case under standard utilities. In particular, two conditions must be satisfied: (i) R must be sufficiently loss averse ($\lambda \geq \lambda^*(p_L, c, \eta)$) and (ii) failures must be sufficiently harmful for the Sender ($L < -\frac{G+p_L W}{1-p_L}$). Intuitively, reference dependence and loss aversion enable R to overcome his commitment problem: thanks to the change in preferences induced by S' initial announcement, R will prefer choosing *Stay* even if the probability of success is low. Since this behavior is costly for S, R behaves *as if* he were punishing S for her lie and this establishes the credibility of S' announcements. However, R is not actually following any punishment motive: his behavior comes from total utility maximization and such utility depends neither on S' utility nor on her intentions.

Proposition 3 states that a message-dependent and fully informative equilibrium may indeed exist; however this equilibrium is not unique. In particular, a message-independent

equilibrium is always possible. This is typical of any communication game: if R ignores S' message, then a strategy in which all messages are sent with a state-independent probability would be optimal for S; this communication strategy, in turn, would justify R's decision to ignore the messages. To be more precise, let $M = \{L, H\}$, and assume S randomizes with equal probability between two constant strategies: $s^L(\cdot) \equiv L$ and $s^H(\cdot) \equiv H$. Denote this strategy with σ^U . Obviously, $\pi(m; \sigma^U) = \frac{1}{2}$ for every $m \in \{L, H\}$. The following proposition holds.

Proposition 4 *Let $\eta > 0$. If assumptions 1, 2 and 3 hold, then there exists a message-independent equilibrium $(\sigma^U, \beta^U, \pi^U)$ such that:*

- $\sigma^U[s^H] = \sigma^U[s^L] = \frac{1}{2}$;
- $\beta^U(\cdot) = (e^U(\cdot), x_L(\cdot), x_H(\cdot))$, where $e^U(m) = 0$, $x_L^U(m) = 0$ and $x_H^U(m) = \chi_H(\lambda)$ for every $m \in \{L, H\}$;²⁶
- $\pi(m; \sigma^U) = \frac{1}{2}$ for every $m \in \{L, H\}$.

Proof. See Appendix 6.1. ■

The equilibrium described in proposition 4 is not the only message-independent equilibrium. Any mixed strategy σ for which $\sum_{s:s(H)=m} \sigma[s] = \sum_{s:s(L)=m} \sigma[s]$ for every message m would lead to a similar result. Furthermore, we could also construct equilibria in which S' announcements contain some information about the state (that is, equilibria in which $M = \{L, H\}$ and $\pi(H; \sigma) > \frac{1}{2}$ and $\pi(L; \sigma) < \frac{1}{2}$), but not enough information to induce R to play *In* (so that $e(m) = 0$ for every message m). All these equilibria lead to the same (message-independent) outcome, namely R choosing *Out*.

3.2 Risky High-Potential Projects

Assumption 3 implies that high-potential projects are risk-free; as a consequence, when R participates in these projects, he can perfectly foresee his final payoff and losses will never arise.²⁷ Suppose instead that $p_H \in (p_L, 1)$. Under this assumption, even if $\pi(m; \sigma) = 1$, any strategy prescribing $e(m) = 1$ and $x_H(m) > 0$ would expose R to potential losses, while playing $e(m) = 0$ would lead to a certain and loss-free payoff. Thus, if loss aversion is high, R's preferred personal strategy may prescribe *Out* even when he is certain that the project has high probability of success.

Our formal analysis starts by noticing that if $\eta = 0$, all equilibria are message-independent; moreover, for any value of $\eta \in (0, 1)$ and $\lambda > 1$, a message-independent equilibrium always exists. These results can be proved following steps similar to those used in propositions 1 and 4. In this Section, we focus on message-dependent equilibria and our characterization is as in section 3.1: first, we assume that S follows communication strategy s^T and we derive R's

²⁶ $\chi_H(\lambda)$ has been defined in proposition 2.

²⁷ Of course losses would still arise if R joins the project thinking that the probability of success is p_H , while the actual probability is only p_L .

preferred personal strategy given s^T , $m \mapsto \beta^T(m)$, where $\beta^T(\cdot) = (e^T(\cdot), x_L^T(\cdot), x_H^T(\cdot))$. Then, taking β^T as given, we show that S has an incentive to follow strategy s^T .

Before beginning our analysis, we define the following thresholds for λ :

$$\begin{aligned}\underline{\lambda}(p_H, p_L, c, \eta) &= 1 + \frac{(c - p_L)(1 + \eta)}{p_H p_L \eta} \\ \bar{\lambda}(p_H, F, c, \eta) &= \max \left\{ 1 + \frac{p_H - F - c}{p_H(1 - p_H)\eta}, \frac{p_H(1 + \eta)}{(F + c)\eta} - \frac{1}{\eta} \right\}\end{aligned}$$

Our analysis will clarify that $\underline{\lambda}(p_H, p_L, c, \eta)$ represents a minimum level of loss aversion: if λ is above this threshold and R played *In* thinking that the project has high probability of success, he will choose *Stay* even if he finds out that the probability of success is only p_L . On the contrary, if λ exceeds $\bar{\lambda}(p_H, F, c, \eta)$, R will play *Out* even if he is certain that the probability of success is p_H .

The following proposition characterizes R's preferred personal strategy given s^T .

Proposition 5 *Let $\eta > 0$. If assumption 1 hold, then the preferred personal strategy given s^T is given by:*

$$\beta^T(L) = (0, 0, \xi_H^L(\lambda)),$$

$$\beta^T(H) = \begin{cases} (1, \xi_L^H(\lambda), 1) & \text{if } \lambda \in (1, \bar{\lambda}(p_H, F, c, \eta)) \\ b \in \{(1, \xi_L^H(\lambda), 1) \cup (0, 0, \xi_H^H(\lambda))\} & \text{if } \lambda = \bar{\lambda}(p_H, F, c, \eta) \\ (0, 0, \xi_H^H(\lambda)) & \text{if } \lambda > \bar{\lambda}(p_H, F, c, \eta) \end{cases},$$

where

$$\xi_H^L(\lambda) = \begin{cases} 0 & \text{if } \lambda > \frac{p_H(1+\eta)}{c\eta} - \frac{1}{\eta} \\ x \in [0, 1] & \text{if } \lambda = \frac{p_H(1+\eta)}{c\eta} - \frac{1}{\eta} \\ 1 & \text{if } \lambda < \frac{p_H(1+\eta)}{c\eta} - \frac{1}{\eta} \end{cases}$$

$$\xi_L^H(\lambda) = \begin{cases} 1 & \text{if } \lambda > \underline{\lambda}(p_H, p_L, c, \eta) \\ x \in [0, 1] & \text{if } \lambda = \underline{\lambda}(p_H, p_L, c, \eta) \\ 0 & \text{if } \lambda < \underline{\lambda}(p_H, p_L, c, \eta) \end{cases}, \quad \xi_H^H(\lambda) = \begin{cases} 0 & \text{if } \lambda > \frac{p_H(1+\eta)}{c\eta} - \frac{1}{\eta} \\ x \in [0, 1] & \text{if } \lambda = \frac{p_H(1+\eta)}{c\eta} - \frac{1}{\eta} \\ 1 & \text{if } \lambda < \frac{p_H(1+\eta)}{c\eta} - \frac{1}{\eta} \end{cases}$$

Proof. See Appendix 6.2. ■

Before describing the equilibrium in more details, we need to comment on the case $\lambda = \bar{\lambda}(p_H, c_1, c_2, \eta)$. In this case, there is a reference-point consistent strategy that prescribes to

play *In* after message H and another reference-point consistent strategy that prescribes to play *Out* after message H . Furthermore, both these strategy induce the same continuation utility after message H . Usually, we would simply assume that R can randomize between these two continuation strategies with any probability; however, with reference dependence the timing of such a randomization matters: if the outcome of the randomizing device is realized before S sends her message, it would not affect the formation of the reference point (which takes place when R receives S ' message). On the contrary, if the randomization happens after S ' announcement, the randomness introduced by the mixed strategy would enter in the formation of the reference point and would complicate the analysis. Proposition 5 overcomes this problem by assuming that R does not mix between these two continuation strategies.

Notice that R 's preferred personal strategy given s^T may prescribe to play *Out* after message m even if $\pi(m; \sigma) = 1$ and assumption 1(ii) implies that consumption utility is maximized by strategy $(1, \cdot, 1)$. This happens because R is loss averse and high-potential projects are no longer risk-free. Indeed, since these projects may fail with some positive probability, R 's participation may result in losses that a loss-averse agent would prefer to avoid. Obviously, the relevance of this loss increases with the coefficient of loss aversion and if λ exceeds $\bar{\lambda}(p_H, F, c, \eta)$, R will prefer to choose *Out* even when he is certain that the project has high probability of success.

In addition to the effect we just described, loss aversion also has the same effect we described in section 3.1: if R joins the project thinking that the probability of success is high and he finds out that this is not the case, loss aversion may induce him to play *Stay* even if this is suboptimal from the consumption-utility point of view. This happens if the coefficient of loss aversion is greater or equal to $\underline{\lambda}(p_H, p_L, c, \eta)$. Notice that, $\underline{\lambda}(1, p_L, c, \eta) = \lambda^*(p_L, c, \eta)$ and if $p_H < 1$, then $\underline{\lambda}(p_H, p_L, c, \eta) > \lambda^*(p_L, c, \eta)$. Intuitively, if $p_H < 1$, R knows that high-potential projects can fail with positive probability; as a consequence, an unexpected early liquidation of the project is less harmful (in terms of gain/loss utility) than in section 3.1 and R is less willing to trade-off consumption utility against a potential decrease in losses.

To sum up, proposition 5 implies that R 's preferred personal strategy given s^T prescribes $(1, 1, 1)$ after message H if and only if the coefficient of loss aversion is neither too high nor too low: $\lambda \in [\underline{\lambda}(p_H, p_L, c, \eta), \bar{\lambda}(p_H, F, c, \eta)]$. The next corollary characterizes the set of parameters for which this range is non-empty.

Corollary 1 *Let $\eta > 0$. If assumption 1 holds, $\underline{\lambda}(p_H, p_L, c, \eta) \leq \bar{\lambda}(p_H, F, c, \eta)$ if and only if $c \leq \bar{c}(p_L, p_H, F, \eta)$, where $\bar{c}(p_L, p_H, F, \eta) > p_L$. Furthermore, the preferred personal strategy given s^T prescribes $(e(H), x_L(H), x_H(H)) = (1, 1, 1)$ at $m = H$ if and only if $\lambda \in [\underline{\lambda}(p_H, p_L, c, \eta), \bar{\lambda}(p_H, F, c, \eta)]$.*

Proof. See Appendix 6.3. ■

Having characterized R 's preferred personal strategy given s^T , we are ready to provide conditions for the existence of a message-dependent and fully informative equilibrium.

Proposition 6 *Let $\eta > 0$. If assumptions 1 and 2 hold, then a message-dependent and fully informative equilibrium exists if and only if $c < \bar{c}(p_L, p_H, F, \eta)$, $\lambda \in [\underline{\lambda}(p_H, p_L, c, \eta), \bar{\lambda}(p_H, F, c, \eta)]$ and $L < -\frac{G+p_L W}{1-p_L}$.*

Proof. See Appendix 6.4. ■

Thus, if we allow high-potential projects to fail with some probability, a message-dependent and fully informative equilibrium still exists, but only under some additional conditions. Whereas proposition 3 requires (i) a sufficiently high degree of loss aversion, and (ii) S to be sufficiently hurt by projects' failure, proposition 6 introduces two additional conditions: (iii) an upper bound on the degree of loss aversion and (iv) an upper bound on the cost associated with the action *Stay*. In particular, if $\lambda > \bar{\lambda}(p_H, F, c, \eta)$, a fully informative equilibrium exists, but it is not message-dependent; in this case, S' inability to affect R's behavior does not come from the lack in commitment power we described before, but from R's reluctance to undertake projects that could result in losses. Thus, a message-dependent and fully informative equilibrium requires $\lambda \in [\underline{\lambda}(p_H, p_L, c, \eta), \bar{\lambda}(p_H, F, c, \eta)]$, which, in turn, requires $c < \bar{c}(p_L, p_H, F, \eta)$.

The following remark summarizes some comparative static results on $\underline{\lambda}(p_H, p_L, c, \eta)$, $\bar{\lambda}(p_H, F, c, \eta)$ and $\bar{c}(p_L, p_H, F, \eta)$.

Remark 4 (i) $\underline{\lambda}(p_H, p_L, c, \eta)$ is increasing in c , decreasing in η , p_H and p_L . Furthermore $\underline{\lambda}(1, p_L, c, \eta) = \lambda^*(p_L, c, \eta)$.

(ii) $\bar{\lambda}(p_H, c, F, \eta)$ is increasing in p_H , decreasing in c , F and η . Furthermore $\bar{\lambda}(p_H, F, c, \eta) \rightarrow \infty$ as $p_H \rightarrow 1$.

(iii) $\bar{c}(p_L, p_H, F, \eta)$ is increasing in p_L . Furthermore, if $p_H = 1$, then $\bar{c}(p_L, p_H, F, \eta) = 1 - F$ and assumption 1 implies $c < \bar{c}(p_L, p_H, F, \eta)$ for every c .

To get the intuition behind these results, notice that an increase in p_H has two effects: (i) high-potential projects succeed more often making R more willing to undertake them ($\bar{\lambda}(p_H, c, F, \eta)$ increases), (ii) if R's preferred personal strategy prescribes to play *In* after message *H*, R's reference point will put a higher weight on outcome *s* and loss aversion will make him less inclined to choose *Liquidate* after finding that the actual probability of success is only p_L ($\underline{\lambda}(p_H, p_L, c, \eta)$ decreases).

Instead, if p_L increases, $\bar{\lambda}(p_H, c, F, \eta)$ stays constant and $\underline{\lambda}(p_H, p_L, c, \eta)$ decreases. The first result follows immediately from the fact that $\bar{\lambda}(p_H, c, F, \eta)$ depends only on R's behavior when the project is high-potential. On the other hand, an increase in p_L makes the decision to play *Stay* when the project is standard more profitable and this lowers the level of loss aversion necessary to induce such a behavior. The latter mechanism also explains why $\underline{\lambda}(p_H, p_L, c, \eta)$ is increasing in c . However, an increase in c also reduces R's willingness to undertake high-potential projects ($\bar{\lambda}(p_H, c, F, \eta)$ decreases).

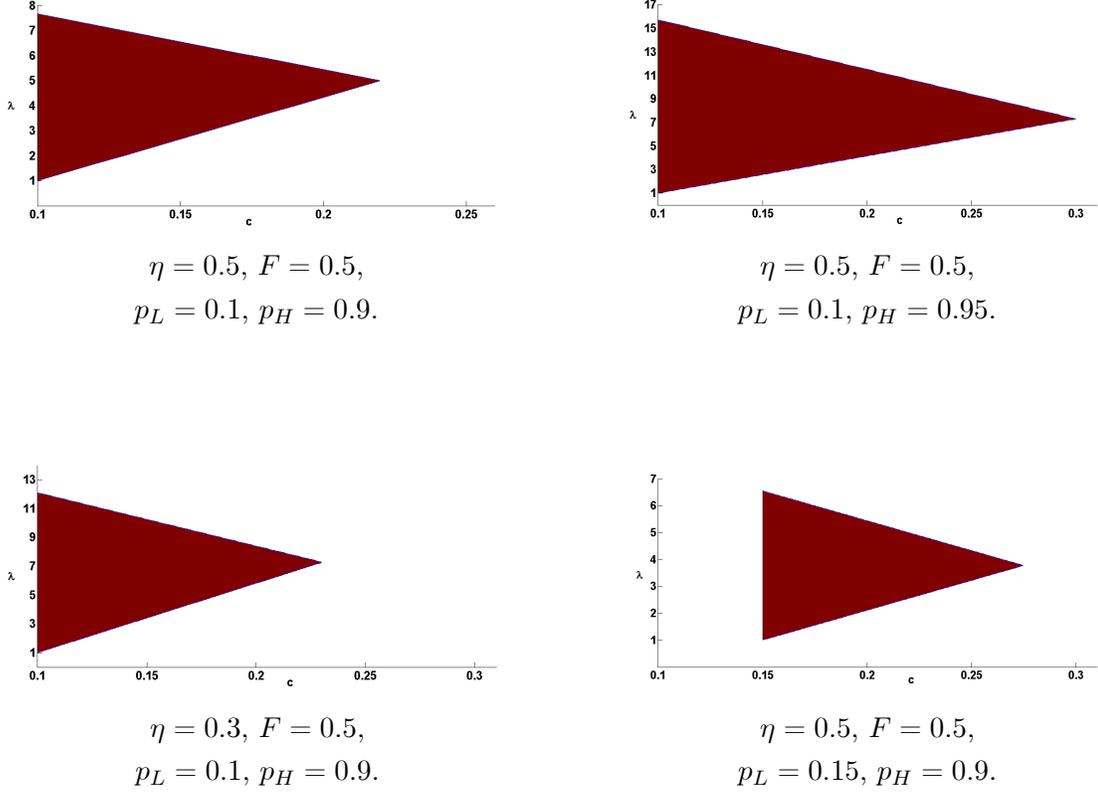


Figure 2: pairs (c, λ) for which a message-dependent and fully informative equilibrium exists

Figure 2 highlights the pairs (c, λ) for which a message-dependent and fully revealing equilibrium exists.²⁸ The increasing linear function, $\underline{\lambda}(p_H, p_L, c, \eta)$, and the decreasing one, $\bar{\lambda}(p_H, F, c, \eta)$, intersect at $\bar{c}(p_L, p_H, F, \eta)$: in line with remark 4, the range of loss aversion coefficients for which a message-dependent and fully informative equilibrium exists shrinks as c increases and becomes empty if $c > \bar{c}(p_L, p_H, F, \eta)$.

3.3 Imperfect Learning of the Project's Quality

The previous results are derived under the assumption that R learns the actual probability of success of the project upon participating in it.

Assume that this is not the case, that is suppose that R learns this probability only with some probability $q \in [0, 1]$. Formally if R plays *In* he receives a signal $\omega \in \{L, \phi, H\}$ according

²⁸In the fourth graph, the blank region comes from assumption 1(i): $c > p_L$.

to the following conditional distribution: for every $Y \in \{L, H\}$

$$\Pr\{\omega = w \mid Y\} = \begin{cases} q & \text{if } w = Y \\ 1 - q & \text{if } w = \phi \\ 0 & \text{if } w = Y', Y' \neq Y \end{cases} \quad (7)$$

Thus, the signal reveals the actual quality of the project with probability q ; with complementary probability, R receives an uninformative message that leaves his beliefs unchanged. For simplicity, suppose assumption 3 holds.²⁹ This game is represented in Figure 3.

Imperfect learning about the project's quality helps supporting a fully informative equilibrium: if R believes S is following a truthful communication strategy, he will believe S' announcement unless he receives a signal that contradicts her. Thus, if S sends message $m = H$ when the project is low quality and q is low, she may hurt herself: with probability $(1 - q)$ R will not learn the truth and will choose to *Stay* thinking that the project will succeed with probability 1. Since this behavior may hurt S, she will prefer to announce the truth from the beginning.

To formalize this discussion, we need to introduce some additional notation. R's strategy is now a profile:

$$\beta(\cdot) = (e(\cdot), x_L(\cdot), x_\phi(\cdot), x_H(\cdot)) : M \rightarrow [0, 1]^4$$

where $x_\omega(m)$ represents the probability with which R plans to choose *Stay* if S sends message m , he plays *In* and he receives signal $\omega \in \{L, \phi, H\}$ ($e(\cdot)$ has the same interpretation as before). In line with our previous notation, $\pi(m, \omega; \sigma)$ is the probability that R assigns to project being high-quality if he believes that S is following communication strategy σ and if he has received message m and signal ω . Thus, recalling that s^T denotes a truthful communication strategy, it is immediate to verify that

$$\pi(m; s^T) = \begin{cases} 1 & \text{if } m = H \\ 0 & \text{if } m = L \end{cases} \quad \text{and} \quad \pi(m, \omega; s^T) = \begin{cases} 1 & \text{if } \omega = H \\ 1 & \text{if } m = H \text{ and } \omega = \phi \\ 0 & \text{otherwise} \end{cases} .$$

The following proposition characterizes message-dependent and fully informative equilibria. As usual, message-independent equilibria also exists, but we will not characterize them.³⁰

²⁹All results could be generalized to the case $p_H \in (0, 1)$.

³⁰Such characterization can be provided following the same steps we used in Proposition 4.

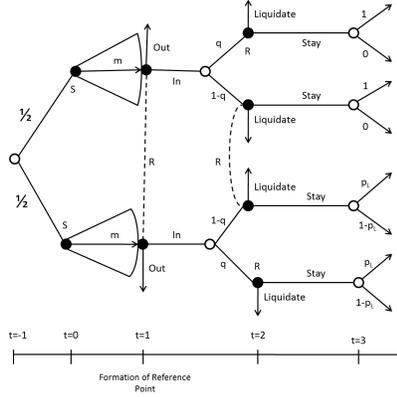


Figure 3: Imperfect Learning of the Project's Quality.

Proposition 7 Let $\eta > 0$. If assumptions 1, 2 and 3 hold and the signaling technology is given by (7), a message-dependent and fully-revealing equilibrium exists if and only if either $\lambda \geq \lambda^*(p_L, c, \eta)$ and $L < -\frac{G+p_L W}{1-p_L}$ or $L < -\left(\frac{G}{(1-q)(1-p_L)} + \frac{p_L W}{(1-p_L)}\right)$ (or both).

Proof. Suppose either $\lambda > \lambda^*(p_L, c, \eta)$ and $L < -\frac{G+p_L W}{1-p_L}$ or $L < -\left(\frac{G}{(1-q)(1-p_L)} + \frac{p_L W}{(1-p_L)}\right)$. If R believes S is following communication strategy s^T , $\pi(H; s^T) = 1$, $\pi(L; s^T) = 0$, $\pi(L, \phi; s^T) = \pi(L, L; s^T) = 0$, $\pi(H, \phi; s^T) = \pi(H, H; s^T) = 1$. Thus, one can adapt the proof of proposition 2 to show that the preferred personal strategy of R given s^T prescribes: $e(H) = 1$, $e(L) = 0$, $x_H(H) = x_\phi(H) = 1$, $x_L(H) = \chi_L(\lambda)$, $x_L(L) = x_\phi(L) = 0$, $x_H(L) = \chi_H(\lambda)$, where $\chi_L(\lambda)$ and $\chi_H(\lambda)$ are defined in proposition 2. In particular, the only new step in the proof is to show that, upon receiving message Y , R 's preferred personal strategy given s^T prescribes the same behavior after signal $\omega = Y$ and $\omega = \phi$ ($Y \in \{L, H\}$). This follows from the fact that after both signals R 's belief about the quality of the project is the same and loss aversion makes playing mixed strategies suboptimal.

Now, consider the communication strategy of player S ; obviously, if the project has high probability of success, she prefers to tell the truth (in this way, she can get $W + G$ instead of 0). Suppose instead that the probability of success is p_L . Then, R 's preferred personal strategy given s^T implies that if she tells the truth, she gets 0, while if she lies, she can get:

$$G + (p_L W + (1 - p_L) L)$$

if $\lambda \geq \lambda^*(p_L, c, \eta)$ and

$$G + (1 - q)(p_L W + (1 - p_L) L)$$

if $\lambda < \lambda^*(p_L, c, \eta)$.³¹ Our assumptions imply that in both cases, S ' utility would be negative.

³¹Once more, we break indifferences by assuming that, whenever indifferent, R plays *Stay* instead of

Thus, she will prefer telling the truth.

Suppose that a message-dependent and fully informative equilibrium exists. R's preferred personal strategy would be the one we described above. Obviously, S will always tell the truth if the project has high probability of success. If the project has low probability of success, S will tell the truth only if:

$$0 \geq G + q \cdot \chi_L(\lambda) \cdot (p_L W + (1 - p_L) L) + (1 - q) \cdot (p_L W + (1 - p_L) L)$$

Thus either $L < -\left(\frac{G}{(1-q)(1-p_L)} + \frac{p_L W}{(1-p_L)}\right)$ or $\lambda \geq \lambda^*(p_L, c, \eta)$ (so that $\chi_L(\lambda) > 0$) and $L < -\left(\frac{G+p_L W}{(1-p_L)}\right)$ (or both). ■

As a special case of the previous proposition, notice that if $q = 0$, a message-dependent and fully-revealing equilibrium exist if and only if $L < -\frac{G+p_L W}{1-p_L}$ and loss aversion does not play any role. Intuitively, if lies cannot be detected, R may play *Stay* when the project has low probability of success simply because he holds incorrect beliefs (induced by S' announcements). In this case, truthtelling will be possible if and only if failures are sufficiently harmful for the Sender. On the other hand, as $q \rightarrow 1$, $\frac{G}{(1-q)(1-p_L)} \rightarrow \infty$ and we are back to the case we characterized in proposition 3. Finally, if $q \in (0, 1)$, a message-dependent and fully revealing equilibrium exists if and only if failures are sufficiently harmful and either R is sufficiently loss averse or he learns the true probability of success with sufficiently low probability.

Proposition 7 states that credible information transmission is achieved because a lie may induce R to keep exerting effort (action *Stay*) on projects that are unlikely to succeed. However the mechanism through which this behavior is achieved depends on whether λ is above or below $\lambda^*(p_L, c, \eta)$. If $\lambda < \lambda^*(p_L, c, \eta)$, R plays *Stay* because he holds the erroneous belief (induced by S' announcements) that the project has high probability of success. Instead, if $\lambda > \lambda^*(p_L, c, \eta)$ R plays *Stay* even if he is fully aware that this action is suboptimal from the consumption utility point of view. Thus, whereas in the former case R's behavior is determined by incorrect beliefs, in the latter one it is due to a change in reference point and to R's desire to avoid losses. As a result, there is a sense in which a truthful equilibrium based on loss aversion is more robust. Indeed, if R's decision to play *Stay* is due to incorrect beliefs, S would have an incentive to send additional messages after that R played *In* in order to change such beliefs. Since at this decision node players' interests are aligned, these messages would be credible and, by eliminating the cost of lying, they would destroy the message-dependent and fully revealing equilibrium. Thus, unless S can commit not to send additional messages after R played *In*, a message-dependent and fully revealing equilibrium will fail to exist. This does not happen with loss aversion; in this case, R's behavior is determined by a change in preferences induced by the change in reference point and additional messages could not bypass this mechanism.

Liquidate after message $m = H$.

4 Extensions and Discussion of the Results

4.1 Message Dependence and Full Information Transmission

Propositions 3 and 6 characterize equilibria in which (i) the message sent by the Sender affects the Receiver's behavior, and (ii) the Sender fully reveals her private information. In this section, we show that the second requirement is without loss of generality; to put it differently, if a message-dependent equilibrium exists, this equilibrium can be assumed to entail full information transmission.

To understand this point, consider a message-dependent equilibrium and let σ and $\beta(\cdot) = (e(\cdot), x_L(\cdot), x_H(\cdot))$ be the equilibrium strategies played by S and R. Then, we can find a message m_1 such that $e(m_1) > 0$. Let $m^* = \arg \max_{m \in M} e(m)$; reference-point consistency and assumption 1(ii) implies $m^* \in \arg \max_{m \in M} \pi(m; \sigma)$, $\pi(m^*; \sigma) > \frac{1}{2}$ and $x_H(m^*) = 1$.³² By assumption 2, this implies that we can assume without loss of generality that $\sigma(s) > 0$ implies $s(H) = m^*$.³³ As a result, for any $m \neq m^*$, $\pi(m; \sigma) = 0$ and $e(m) = 0$ (this last implication follows from reference point consistency and assumption 1). Since $\pi(m^*; \sigma) > \frac{1}{2}$, S must send some message $m \neq m^*$ in state L with positive probability and her utility must be equal to 0. Thus, without loss of generality, we can conclude that $\sigma(s) > 0$ implies $s(L) = m^{**} \neq m^*$.

4.2 Reference Point and Mental Accounting

In the paper we assumed R is subject to mental accounting and evaluates payoffs as if they belonged to one of two different categories: early and late payoffs. As a consequence, gains and losses are evaluated separately for each of these two dimensions.

Although mental accounting is a rather standard assumption in the literature on reference dependence, this section shows that our results do not hinge on it. In particular, if we remove the assumption of piecewise linearity of function $\mu(\cdot)$, namely the function capturing gain/loss utility, the insight of the paper would go through even if R did not distinguish among different payoff's dimensions. To show this, we consider the case in which $p_H = 1$ and R perfectly learns the project's actual probability of success upon playing In .

Since $\mu(\cdot)$ is given by (3), it satisfies the following properties:³⁴

P1 regularity: $\mu(0) = 0$ and $\mu(\cdot)$ is continuous and twice differentiable for $x \neq 0$;

P2 strict monotonicity: $\mu(\cdot)$ is strictly increasing;

P3 loss aversion for large stakes: for any $x, y \in \mathbb{R}$ such that $x > y \geq 0$, $\mu(x) + \mu(-x) < \mu(y) + \mu(-y)$;

³²These results can be proved following the same steps used in Propositions 3 and 6.

³³Obviously, S can also send other messages, but all of them would lead R to play In with probability $e(m^*)$. In this respect, all these messages can be considered to be the same.

³⁴These properties are taken from Koszegi and Rabin, 2006, 2007, 2009.

P4 loss aversion for small stakes: $\frac{\lim_{x \rightarrow 0} \mu'(-|x|)}{\lim_{x \rightarrow 0} \mu'(|x|)} = \lambda > 1$;

P5 constant sensitivity: $\mu''(x) = 0$ for any $x \in \mathbb{R}$;

Let P5 be replaced with the following property:

P5' diminishing sensitivity: $\mu''(x) < 0$ if $x > 0$ and $\mu''(x) > 0$ if $x < 0$.

P5' states that the marginal importance of gain/loss utility decreases as the distance between the actual utility and the reference utility increases. Assume R evaluates payoffs as if they belonged to a unique dimension. The next proposition shows that if $\mu(\cdot)$ satisfies properties P1-P4 and property P5', we can provide conditions under which R's preferred personal strategy given s^T prescribes to play *Stay* even after finding out that the project's probability of success is p_L .

Proposition 8 *Let $\mu(\cdot)$ satisfy properties P1-P4 and P5'. If assumptions 1 and 3 hold, then R's preferred personal strategy prescribes $e^T(L) = 0$, $e^T(H) = 1$ and $x_H^T(H) = 1$ and³⁵*

$$x_L^T(H) = \begin{cases} 1 & \text{if } p_L - c + (1 - p_L) \mu(-1) > \mu(-1 + c) \\ v \in [0, 1] & \text{if } p_L - c + (1 - p_L) \mu(-1) = \mu(-1 + c) \\ 0 & \text{if } p_L - c + (1 - p_L) \mu(-1) < \mu(-1 + c) \end{cases} .$$

Proof. See Appendix 6.5. ■

Thus, if $p_L - c + (1 - p_L) \mu(-1) > \mu(-1 + c)$, the preferred personal strategy given s^T will have R playing *Stay* after message $m = H$ even if he discovers that the project has small probability of success. Then, we can replicate the proof of proposition 3 and conclude that a message-dependent and fully informative equilibrium exists if and only if:

$$L < -\frac{G + p_L W}{1 - p_L}$$

and:

$$p_L - c + (1 - p_L) \mu(-1) > \mu(-1 + c). \quad (8)$$

It is immediate to see that (8) can be rewritten as:

$$\int_{-1}^{-1+c} \mu'(x) dx < p_L - c - p_L \left(\int_{-1}^0 \mu'(x) dx \right)$$

³⁵The characterization of $x_L^T(L)$ and $x_H^T(L)$ is provided in the proof of the proposition.

If P5 holds (for instance because $\mu(\cdot)$ is given by (3)), (8) would require $p_L > c$ contradicting assumption 1. Therefore, the existence of a message-dependent and fully informative equilibrium requires the relevance of the gain/loss utility to be sufficiently higher in the interval $(-1 + c, 0)$ than in the interval $(-1, -1 + c)$.

4.3 Credibility and Monetary Transfers

In section 3 we assume that non-binding announcements are the only tool S can use to transmit her private information to R. However, in some environments these announcements can be backed by *enforceable monetary promises*. This section studies how this tool can complement loss aversion and help S affecting R's behavior.

Suppose $p_H < 1$ and consider the game in section 3.2. Assume that at the initial node S can not only send a message m , but that he can also promise a monetary transfer to R if certain contingencies occur. To facilitate the comparison with the previous section we assume that monetary promises can be verified and enforced at no cost by a third party (for instance, a judge or a mediator) and that S does not incur any direct cost in writing these enforceable clauses. Monetary transfers can be represented as functions $\kappa : \mathcal{C} \rightarrow \mathbb{R}_+$, where \mathcal{C} represents a set of verifiable contingencies, namely a partition of the set of histories representing the information sets that the third party can verify. We will assume that monetary transfers enter additively in players' utility and that R accounts them in the late payoff dimension.³⁶

In our model these monetary transfers may serve two purposes. On the one hand, if S knows that the project is high-potential, she can use transfers to establish her credibility by promising to transfers money after contingencies that are more likely if the project has a low probability of success; we will refer to this objective as to the *credibility motive*. On the other hand, S can use transfers to insure R against possible losses overcoming in this way his loss aversion and inducing him to play *In* even though, absent any monetary disbursement, he would have played *Out*; we refer to this as to the *participation motive*. To simplify the discussion, we assume that the participation motive is relevant only when the project is high-potential. Formally, this can be done assuming that if the project has low-probability of success S does not want to transfer money to R in order to achieve his participation in the project, that is $G < F$ and $L < -\frac{G+p_L W}{1-p_L}$.³⁷

Under these assumptions, we can reinterpret the analysis in section 3.2 as follows. If monetary transfers are not allowed, the credibility motive can be achieved if and only if $\lambda \geq \underline{\lambda}(p_H, p_L, c, \eta)$, while the participation motive can be attained if and only if $\lambda \leq \bar{\lambda}(p_H, F, c, \eta)$. As a consequence, if $c \leq \bar{c}(p_L, p_H, F, \eta)$ and $\lambda \in [\underline{\lambda}(p_H, p_L, c, \eta), \bar{\lambda}(p_H, F, c, \eta)]$, both these motives can be achieved without transfers. Therefore, we will focus on the case in which

³⁶This last assumption can be easily generalized.

³⁷If $p_H = 1$, the participation motive is not relevant and credibility can be gained for free by the sender by promising to transfer money if the the outcome fails (a contingencies that never arises if the actual probability of success is equal to p_H)

$c \leq \bar{c}(p_L, p_H, F, \eta)$, but $\lambda \notin [\underline{\lambda}(p_H, p_L, c, \eta), \bar{\lambda}(p_H, F, c, \eta)]$. Two cases are possible.

If $\lambda > \bar{\lambda}(p_H, F, c, \eta)$, the credibility motive can be achieved using non-binding announcements, while the participation one cannot. To fix this problem, S can promise money if the project fails, helping him to overcome his loss aversion. It is not hard to see that the amount S has to promise R is equal to 0 when $\lambda = \bar{\lambda}(p_H, F, c, \eta)$ and is linearly increasing in λ .³⁸ Since S' payoff is bounded, there exists a value of loss aversion, say λ^+ , above which achieving R's participation is too costly and S prefers not to do that.

Suppose instead that $\lambda < \underline{\lambda}(p_H, p_L, c, \eta)$. In this case, S must use monetary transfers to establish her credibility, but once she has done that, R's participation will come for free. Obviously, if the actual probability of success is verifiable, credibility can be attained by promising to pay R if the probability of success is different from what S announced at the beginning of the game. Thus, we will focus on the more interesting case in which probabilities p_H and p_L are not verifiable. In this case, credibility can be established either by promising to pay R if he decides to play *Liquidate* or by promising him money if he plays *Stay* and the project fails.³⁹ However, the first type of promises could establish S' credibility only if she can credibly commit to get negative utility (recall we are assuming $G < F$). On the other hand, since $p_H < 1$, the second type of promises are costly for the Sender; in particular, the monetary promises needed in this case would be equal to 0 if $\lambda = \underline{\lambda}(p_H, p_L, c, \eta)$ and would be linearly decreasing in λ for values of λ lower than $\underline{\lambda}(p_H, p_L, c, \eta)$.⁴⁰ As a result, we can find a threshold, λ^- (possibly lower than 1), below which attaining credibility would be too costly and the Sender prefers not to do that.

To summarize, the interaction between loss aversion, participation motive and credibility motive could lead to a non-monotonic relation between loss aversion (measured by coefficient λ) and the amount of monetary transfers S has to promise in order to induce R's participation in high-quality projects; in particular, this amount is positive (and possibly decreasing) in the interval $(\lambda^-, \underline{\lambda}(p_H, p_L, c, \eta))$, equal to 0 in the interval $[\underline{\lambda}(p_H, p_L, c, \eta), \bar{\lambda}(p_H, F, c, \eta)]$ and positive and increasing in λ if $\lambda \in (\bar{\lambda}(p_H, F, c, \eta), \lambda^+)$ and $p_H < 1$.

4.4 Sender with Reference Dependent Preferences

Throughout the paper, we assume that R is the only agent who exhibits reference-dependent preferences. If we want to introduce reference dependence and loss aversion in S' preferences, it is natural to assume that her reference point is determined before she sends the message, that is when she knows the true quality of the project and can use strategic reasoning to conjecture how the game will be played. In our model, S' loss aversion would discourage her

³⁸The constant marginal increase is a consequence of our assumption about the linear structure of the gain/loss utility.

³⁹Indeed, conditional on R playing *In*, both these contingencies are more likely to happen when the project has low probability of success.

⁴⁰Linearity holds for the same reason as before, but the absolute value of the slope would be different than the one for the case $\lambda > \bar{\lambda}(p_H, F, c, \eta)$.

to take actions that could result in losses. Therefore, we can distinguish between two cases depending on whether assumption 3 holds or not.

If $p_H = 1$, in equilibrium, losses arises only if R plays *Stay* even when the probability of success is low.⁴¹ Replicating the analysis of section 3.1, we can conclude that propositions 2 and 3 still hold and S' reference dependence does not affect the main insight of the paper.

Instead, if $p_H < 1$, Sender's loss aversion may be relevant. To see why, let η_i and λ_i be the parameters describing i 's total utility ($i \in \{S, R\}$). Proposition 5 still holds and implies that if $\lambda_R \in [\underline{\lambda}(p_H, p_L, c, \eta_R), \bar{\lambda}(p_H, F, c, \eta_R)]$, S could induce participation by sending message $m = H$. However, this message exposes S to some potential losses and her total utility would be given by:

$$W + p_H G + (1 - p_H) L - p_H (1 - p_H) \eta_S (\lambda_S - 1) (G - L)$$

Therefore, one can show that the existence of a message-dependent and fully revealing equilibrium requires:

$$\lambda_S \leq 1 + \frac{W + p_H G + (1 - p_H) L}{p_H (1 - p_H) \eta_S (G - L)}$$

If λ_S is above this threshold, the only equilibrium will be message-independent; intuitively, if S is sufficiently loss averse, she will prefer not to expose herself to potential losses even when she knows that the project is likely to succeed. As a result, she will prefer sending uninformative messages and this will induce R to play *Out*.

4.5 Reference Point's Formation and Updating

This paper assumes that R's reference point is determined once and for all upon listening to the Sender's announcement. This assumption has important implications for our analysis.

On the one hand, the thought process through which R's reference point is established takes place when he has to choose between *In* and *Out* and not before S made her announcement. This simplifies the analysis and is also sensible: since in a rational expectation approach à la Köszegi and Rabin the formation of the reference point requires R to conjecture how the game will evolve, it seems plausible to assume that R will undergo this mental process only when he has to make his initial choice.

On the other hand, our assumption concerning the timing of the reference point's formation also implies that R will not adjust the reference point after finding out the true quality of the project. Although this approach maximizes the influence of S' messages on R's reference point, our results do not crucially hinge on it: as long as S' initial announcement plays a sufficiently important and irreversible role in the formation of R's reference point, our qualitative findings would still hold. For instance, we could assume that the reference point is determined according to the following process: after S' announcement the reference point

⁴¹In principle, S' reference dependence could also matter if R's preferred personal strategy prescribed randomization at some decision node. However, since the analysis of sections 3.1 and 3.2 still holds, this will never be the case.

is determined as in section 2.3, but upon learning the actual probability of success, the probability that the reference point assigns to each outcome is equal to a weighted average between the old reference point and a new distribution based on an optimizing behavior given all available information.⁴² As long as the weight put on the old reference point is positive and sufficiently large, reference dependence could still induce S' truthtelling through a mechanism similar to the one described in Section 3.

4.6 Reference Dependence and Truthtelling

Section 3 shows that reference dependence and loss aversion may play an important role in the transmission of information; the reason is that S' announcement can modify R's reference point and, through this channel, his long run behavior. In particular, reference dependence gives saliency to the payoff that would arise on (what R believes is) the equilibrium path; if R is surprised by unexpected events (such as, those arising after a lie when R was trusting S), loss aversion may lead him to deviate from the consumption-utility-maximizing behavior. In our model, this change in R's long-run behavior helps aligning his interests with S' ones and enables information transmission in an environment in which this would not be otherwise possible.

Thus, a natural question arises: can reference dependence and loss aversion prevent information transmission instead of facilitating it? The answer to this question is affirmative as long as the change in reference point induced by S' announcement modifies R's long run behavior and introduces a conflict of interests between R and S that was previously absent. For instance, consider a model similar to the one we have analyzed in which there are three types of projects: those that fail for sure as soon as R participates in it (low-quality projects), those that succeed with probability $p < 1$ (medium-quality projects) and those that succeed with probability 1 (high-quality projects).⁴³ Assume that high- and low-quality projects have *ex-ante* probability $\frac{1-\varepsilon}{2}$, while medium-quality projects have probability $\varepsilon \in (0, 1)$; finally suppose that L is negative and large in absolute value and that the Sender knows whether a project is low quality or not, but cannot distinguish between medium- and high-quality ones.

Under standard utility, if ε is small, R will play *In* whenever he is sufficiently certain that the project is not low-quality. Since the failure of low-quality projects is harmful for both players, S will reveal her information truthfully and R will trust her. Now, suppose that R has reference-dependent preferences and is loss averse; then the news that the project is not low-quality together with the assumption that ε is small give saliency to the outcomes that occur when the project is high-quality and this may induce R to play *Stay* even when the project is medium quality and p is relatively small.⁴⁴ Since this type of behavior may hurt S

⁴²Obviously, in this case the definition of reference-point consistency and preferred personal strategies should be adapted to take into account this updating.

⁴³Thus, if R plays *In* when the project is low-quality, he does not have to choose between *Liquidate* and *Stay*, while for the other two types of projects this decision node is present.

⁴⁴The intuition behind this phenomenon is similar to the one we gave in Section 3.

(L is negative and large in absolute value), she may prefer not to reveal her information and to give up the possibility of inducing participation even when she knows that the project is not low-quality.

5 Conclusion

In this paper we considered a model of communication in which the receiver has reference-dependent preferences and is loss averse. Under these assumptions, the announcement made by the sender can affect the receiver's reference point and, consequently, his optimal behavior. Using this insight, we showed that reference dependence and loss aversion can help the sender establishing credible information transmission and affecting receiver's behavior.

In particular, we consider a model in which the receiver has to decide whether to undertake a project whose probability of success is initially known only to the sender. Although under standard utility the conflict of interest between the two parties destroys the credibility of the sender and prevents her from affecting receiver's behavior with her announcements, reference dependence and loss aversion in receiver's preferences may overcome this problem. This happens if the coefficient of loss aversion does not take extreme values: it must be sufficiently high to affect the receiver's behavior, but it should not be too high because otherwise the receiver would never undertake risky projects.

The simple structure of the game we analyzed in this paper should allow for an experimental design to test whether credible communication could affect the behavior of the receiver through the change in his reference point, and whether this mechanism could help establishing credible information transmission.

More in general, a growing literature investigates the importance of psychological phenomena in the analysis of strategic interactions.⁴⁵ In our opinion, communication games represent a natural benchmark to study the extent to which an agent can affect his opponents' psychology, emotional status or expectations. Therefore, a better understanding of the link between communication and behavioral biases is highly desirable both from the theoretical and the experimental point of view.

6 Appendix

6.1 Proof of Proposition 4

If R behaves as described in the proposition, σ^U is optimal for S and the beliefs are the ones described in the statement of the proposition. Thus, we only need to show that β^U is the preferred personal strategy given σ^U .

⁴⁵See section 1.1 for a partial overview of this literature.

In this message-independent equilibrium, R does not modify his beliefs concerning the state of nature. Therefore, R assigns probability $\frac{1}{2}$ to each type of project after any message $m \in M$. Since the message does not modify R's belief, we will not specify it in our characterization and we will assume that the continuation strategy after message m is the same for any message m . Therefore, the continuation strategy after message m associated with R's preferred personal strategy will be denoted with $\beta = (e, x_L, x_H)$. If R follows this strategy, his reference utility would be:

$$\tilde{u} \left(\tilde{\zeta}_M (\beta; \sigma^U) \right) = \left(\tilde{u}_1 \left(\tilde{\zeta}_M (\beta; \sigma^U) \right), \tilde{u}_2 \left(\tilde{\zeta}_M (\beta; \sigma^U) \right) \right)$$

where:

$$\tilde{u}_1 \left(\tilde{\zeta}_M (\beta; \sigma^U) \right) [w] = \begin{cases} 1 - e & \text{if } w = 0 \\ \frac{e}{2} \cdot (2 - x_H - x_L) & \text{if } w = -F \\ \frac{e}{2} \cdot (x_H + x_L) & \text{if } w = -F - c \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

in the first dimension and

$$\tilde{u}_2 \left(\tilde{\zeta}_M (\beta; \sigma^U) \right) [w] = \begin{cases} \left(\frac{e \cdot x_H}{2} + \frac{e \cdot x_L \cdot p_L}{2} \right) & \text{if } w = 1 \\ 1 - \left(\frac{e \cdot x_H}{2} + \frac{e \cdot x_L \cdot p_L}{2} \right) & \text{if } w = 0 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

in the second dimension.

Let $\beta^U = (0, x_L, x_H)$, $x_L, x_H \in [0, 1]$.⁴⁶ In this case $\nu \left(\tilde{\zeta}_M (\beta^U; \sigma^U) \mid \tilde{\zeta}_M (\beta^U; \sigma^U) \right) = 0$. Consider a deviation $\beta' = (e', x'_L, x'_H)$ with $e' > 0$. Keeping the reference point fixed at $\tilde{\zeta}_M (\beta^U; \sigma^U)$, R's total utility would be given by:

$$e' \cdot \left(\left(\frac{x'_L \cdot p_L + x'_H}{2} \right) (1 + \eta) - F (1 + \eta \lambda) - c \left(\frac{x'_L + x'_H}{2} \right) (1 + \eta \lambda) \right) < 0$$

where the inequality follows from assumptions 1(i) and (iii). Thus, it is immediate to check that β^U is a reference-point consistent strategy if and only if $x_L = 0$ and $x_H = \chi_H(\lambda)$. We conclude that $\beta^U = (0, 0, \chi_H(\lambda))$ is the unique reference-point consistent strategy given σ^U prescribing $e = 0$.

Now, we will show that any other reference-point consistent strategy leads to a total utility lower than 0. First, consider continuation strategies where $\beta = (1, x_L, x_H)$. It is immediate to see that if $x_L = x_H = 0$, this cannot be part of a preferred personal strategy since since the

⁴⁶Given that $e = 0$, x_L and x_H do not affect the formation of the reference point.

utility associated with this strategy is $-F < 0$. Furthermore, if the continuation strategies prescribes $e = 1$, reference-point consistency requires $x_H = 1$ whenever $x_L > 0$. To see this, observe that reference-point consistency and $x_L > 0$ require:

$$p_L - c - \eta\lambda \left(\frac{2 - x_H - x_L}{2} \right) c - \eta\lambda \left(\frac{x_H + x_L p_L}{2} \right) (1 - p_L) + \\ + \eta \left(\frac{2 - x_H - x_L p_L}{2} \right) p_L \geq \eta \left(\frac{x_H + x_L}{2} \right) c - \eta\lambda \left(\frac{x_H + x_L p_L}{2} \right),$$

which implies:

$$1 - c - \eta\lambda \left(\frac{2 - x_H - x_L}{2} \right) c + \left(\frac{2 - x_H - p_L x_L}{2} \right) \eta > \eta \left(\frac{x_H + x_L}{2} \right) c - \eta\lambda \left(\frac{x_H + x_L p_L}{2} \right)$$

and, consequently, $x_H = 1$. Let $\beta = (1, x_L, 1)$ with $x_L \in (0, 1]$. The total utility associated with this strategy would be:

$$\frac{1 + x_L p_L}{2} - F - \frac{1 + x_L}{2} c - \left(\frac{1 - p_L^2 x_L^2}{2} \right) \eta (\lambda - 1) - \frac{1 - x_L^2}{4} c \eta (\lambda - 1)$$

which is negative by assumptions 1(i)-(iii). Finally, consider strategy $\beta = (1, 0, x_H)$ with $x_H \in (0, 1]$; the utility associated with this strategy would be:

$$\frac{1}{2} - F - \frac{c}{2} - \frac{x_H}{2} \left(1 - \frac{x_H}{2} \right) \eta (\lambda - 1) - \left(1 - \frac{x_H}{2} \right) \frac{x_H}{2} c \eta (\lambda - 1).$$

Once more, this expression is negative by assumptions 1(i)-(iii). We conclude that there is no preferred personal strategy given σ^U with $e = 1$.

Now suppose that R plans to follow strategy $\beta = (e, x_L, x_H)$ with $e \in (0, 1)$. Reasoning as before, we can show that reference-point consistency implies $x_H = 1$ whenever $x_L > 0$. Furthermore, R's total utility if $x_L > 0$ would be given by:

$$\frac{1 + x_L p_L}{2} - F - c \left(\frac{1 + x_L}{2} \right) - \eta\lambda \left(\frac{e \cdot (1 + x_L p_L)}{2} \right) \left(\frac{1 - p_L x_L}{2} \right) - \frac{e(1 - x_L^2)}{4} \eta (\lambda - 1) c + \\ + \left(1 - \frac{e \cdot (1 + x_L p_L)}{2} \right) \left(\frac{1 + p_L x_L}{2} \right) \eta - (1 - e) \left(F + \left(\frac{1 + x_L}{2} \right) c \right) \eta\lambda$$

This expression is decreasing in λ , so it is lower or equal than:

$$\left(\frac{1 + x_L p_L}{2} - F - c \left(\frac{1 + x_L}{2} \right) \right) (1 + \eta(1 - e))$$

which is negative by assumptions 1(i) and 1(iii). Thus, a preferred personal strategy given σ^U cannot prescribe a continuation strategy in which $e \in (0, 1)$ and $x_L > 0$. Now consider

strategies β in which $e \in (0, 1)$ and $x_L = 0$. $(e, 0, 0)$ is not the preferred personal strategy given σ^U since total utility would be equal to $-eF - \eta(\lambda - 1)e(1 - e)F < 0$. Then we need $x_H \in (0, 1]$. In this case, total utility would be equal to:

$$\begin{aligned} \frac{x_H}{2} - F - \frac{x_{Hc}}{2} - \eta\lambda \frac{e \cdot x_H}{2} \left(1 - \frac{x_H}{2}\right) + \eta \left(1 - \frac{e \cdot x_H}{2}\right) \frac{x_H}{2} - \\ - \eta\lambda(1 - e) \left(\left(1 - \frac{x_H}{2}\right) F + \frac{x_H}{2} (F + c) \right) - \left(1 - \frac{x_H}{2}\right) \left(\frac{e \cdot x_H}{2}\right) c\eta(\lambda - 1) \end{aligned}$$

Once more, setting $\lambda = 1$ and using assumptions 1(i)-(iii), we can conclude that this utility is negative. Thus, there is no preferred personal strategy in which $e \in (0, 1)$.

Therefore, the preferred personal strategy given σ^U prescribes $\beta^U(m) = (0, 0, \chi_H(\lambda))$ after any message m .

6.2 Proof of Proposition 5

Throughout this proof we assume that S follows communication strategy s^T . Therefore, we omit to specify the dependence of strategies and beliefs on it.

Consider first the case in which S sent message $m = L$. Since $\pi(L) = 0$, $x_H(L)$ is irrelevant in determining the reference point (and reference utility). Then we can follow the same steps in the proof of Proposition 2 and conclude that the preferred personal strategy given s^T is such that $\beta^T(L) = (0, 0, \xi_L^H(\lambda))$, where:

$$\xi_L^H(\lambda) = \begin{cases} 0 & \text{if } \lambda > \frac{p_H(1+\eta)}{c\eta} - \frac{1}{\eta} \\ x \in [0, 1] & \text{if } \lambda = \frac{p_H(1+\eta)}{c\eta} - \frac{1}{\eta} \\ 1 & \text{if } \lambda < \frac{p_H(1+\eta)}{c\eta} - \frac{1}{\eta} \end{cases}$$

Now, suppose S sent message $m = H$. Since $\pi(H) = 1$, $x_L(H)$ is irrelevant in determining R's reference point (and reference utility). Suppose that R is following strategy β such that $\beta(H) = (1, x_L(H), 1)$. In this case R's reference utility would be given by:

$$\begin{aligned} \tilde{u}_1 \left(\tilde{\zeta}_M(\beta; H) \right) [w] &= \begin{cases} 1 & \text{if } w = -c - F \\ 0 & \text{otherwise} \end{cases}, \\ \tilde{u}_2 \left(\tilde{\zeta}_M(\beta; H) \right) [w] &= \begin{cases} p_H & \text{if } w = 1 \\ 1 - p_H & \text{if } w = 0 \\ 0 & \text{otherwise} \end{cases}. \end{aligned}$$

Thus $\nu\left(\tilde{\zeta}_M(\beta; s^T) \mid \tilde{\zeta}_M(\beta; s^T)\right) = p_H - F - c - p_H(1 - p_H)\eta(\lambda - 1)$. β is reference point consistent at $m = H$ only if the following two conditions are satisfied:

$$\begin{aligned} p_H - F - c - p_H(1 - p_H)\eta(\lambda - 1) &\geq \eta(F + c) - \eta\lambda p_H, \\ p_H - F - c - p_H(1 - p_H)\eta(\lambda - 1) &\geq -F + \eta c - \eta\lambda p_H. \end{aligned}$$

By assumption 1(ii), one can easily verify that both inequalities are satisfied. Furthermore if R finds out that the project has low probability of success, his utility from playing *Stay* would be given by $p_L - F - c + \eta p_L(1 - p_H) - \eta\lambda(1 - p_L)p_H$, while his utility from playing *Liquidate* would be given by $-F + \eta c - \eta\lambda p_H$. Comparing these two expressions, we conclude that reference-point consistency requires $x_L(H) = \xi_L^H(\lambda)$ and that $(1, \xi_L^H(\lambda), 1)$ after message $m = H$ is reference-point consistent at $m = H$ given s^T .

$$p_H - p_H(1 - p_H)\eta(\lambda - 1) + \eta\lambda p_H \geq (1 + \eta)(F + c).$$

Now suppose that the preferred personal strategy β prescribes any other continuation strategy $(e(H), x_L(H), x_H(H))$ after message $m = H$. To simplify the notation we will write (e, x_L, x_H) instead of $(e(H), x_L(H), x_H(H))$. The reference point associated with this strategy after message H , would be given by:

$$\tilde{u}_1\left(\tilde{\zeta}_M(\beta; H)\right)[w] = \begin{cases} 1 - e & \text{if } x = 0 \\ e \cdot (1 - x_H) & \text{if } w = -F \\ e \cdot x_H & \text{if } w = -F - c \\ 0 & \text{otherwise} \end{cases},$$

$$\tilde{u}_2\left(\tilde{\zeta}_M(\beta; H)\right)[w] = \begin{cases} e \cdot x_H \cdot p_H & \text{if } w = 1 \\ 1 - e \cdot x_H \cdot p_H & \text{if } w = 0 \\ 0 & \text{otherwise} \end{cases}.$$

We can have several cases. Suppose $e = 1$ and $x_H \in [0, 1)$ (if $x_H = 1$, we already know that the unique reference point consistent strategy would prescribe $(1, \xi_L^H(\lambda), 1)$). $\beta(\cdot)$ will be reference-point consistent at $m = H$ given s^T as long as:

$$p_H x_H - F - c x_H - \eta(\lambda - 1) x_H p_H (1 - x_H p_H) - \eta(\lambda - 1)(1 - x_H) x_H c \geq \eta F + \eta c x_H - \eta\lambda x_H p_H \quad (11)$$

and

$$p_H - c + \eta\lambda x_H p_H^2 + \eta p_H (1 - x_H p_H) - \eta\lambda (1 - x_H) c \leq \eta x_H c \quad (12)$$

Since (11) can be rewritten as:

$$x_H (p_H - c + \eta\lambda x_H p_H^2 + \eta p_H (1 - x_H p_H) - \eta\lambda (1 - x_H) c - \eta x_H c) \geq (1 + \eta) F,$$

we can use (12) to verify that (11) cannot be satisfied.

Now suppose $e = 0$. This strategy is reference-point consistent if and only if (i) $x_L = 0$,

$$(ii) 0 \geq p_H x'_H - F - x'_H c - \eta\lambda (F + x'_H c) + p_H x'_H \eta \text{ for every } x'_H \in [0, 1] \quad (13)$$

and (iii) $x_H = \xi_H^H(\lambda)$. The right hand side of inequality (13) is maximized at $x'_H = 0$ if $\lambda > \frac{p_H(1+\eta)}{c\eta} - \frac{1}{\eta}$ and at $x'_H = 1$ if $\lambda < \frac{p_H(1+\eta)}{c\eta} - \frac{1}{\eta}$. Furthermore, if $x'_H = 0$, inequality (13) is always satisfied, while if $x_H = 1$, inequality 13 holds if and only if $\lambda \geq \frac{p_H(1+\eta)}{(F+c)\eta} - \frac{1}{\eta}$. Since, $\frac{p_H(1+\eta)}{(F+c)\eta} - \frac{1}{\eta} < \frac{p_H(1+\eta)}{c\eta} - \frac{1}{\eta}$, we conclude that if $\lambda \geq \frac{p_H(1+\eta)}{(F+c)\eta} - \frac{1}{\eta}$, there is a reference-point consistent strategy given s^T prescribing $(0, 0, \xi_H^H(\lambda))$ after message $m = H$ (and this is the unique reference-point consistent strategy that prescribes $e = 0$ after message $m = H$).

Finally suppose $e \in (0, 1)$. If $x_H = 1$, we can have a reference-point consistent strategy only if:

$$\begin{aligned} -\eta\lambda e p_H + \eta e (F + c) &= p_H - F - c + \\ &+ \eta(1 - e p_H) p_H - \eta\lambda e p_H (1 - p_H) - \eta\lambda (1 - e) (F + c) \end{aligned}$$

or equivalently:

$$\lambda = \frac{(F + c - p_H - \eta p_H + e\eta p_H^2 + (F + c) e\eta)}{e\eta p_H^2 - (F + c)(1 - e)\eta}$$

Let $\tilde{\lambda}(p_H, F, c, \eta, e) = \frac{(F+c-p_H-\eta p_H+e\eta p_H^2+(F+c)e\eta)}{e\eta p_H^2-(F+c)(1-e)\eta}$. Assumption 1(ii) implies that $\frac{\partial \tilde{\lambda}(p_H, F, c, \eta, e)}{\partial e} > 0$ for any values of the parameters and $\tilde{\lambda}(p_H, F, c, \eta, 1) = 1 - (1 + \eta) \frac{p_H - (F+c)}{\eta p_H^2} < 1$; we conclude that this strategy cannot be reference-point consistent at $m = H$ given s^T . If instead $e \in (0, 1)$ and $x_H < 1$, reference point consistency requires:

$$\begin{aligned} -\eta\lambda e x_H p_H + \eta e F + \eta e x_H c &= x_H p_H - F - x_H c + \eta x_H p_H (1 - e x_H p_H) \\ &- \eta\lambda (1 - x_H p_H) e x_H p_H - \eta\lambda (1 - e) F - \eta\lambda x_H (1 - e x_H) c + \eta(1 - x_H) e x_H c \end{aligned}$$

and at the same time:

$$p_H - c + \eta p_H (1 - e x_H p_H) - \eta\lambda (1 - p_H) e x_H p_H - \eta\lambda (1 - e x_H) c \leq -\eta\lambda e x_H p_H + \eta e x_H c$$

The second inequality can be rewritten as:

$$p_H - c + \eta p_H (1 - ex_H p_H) - \eta \lambda (1 - ex_H) c \leq \eta ex_H c - \eta \lambda ex_H p_H^2$$

Using this expression in the first equality and simplifying, we get:

$$\eta e \leq -1 - \eta \lambda (1 - e)$$

which cannot hold. We conclude that there is no reference-point consistent strategy given s^T that prescribes $e \in (0, 1)$ after message $m = H$.

We can summarize the previous analysis as follows: if $\lambda < \frac{p_H(1+\eta)}{(F+c)\eta} - \frac{1}{\eta}$, the only reference-point consistent strategy at $m = H$ given s^T is $(1, \xi_L^H(\lambda), 1)$. If $\lambda \geq \frac{p_H(1+\eta)}{(F+c)\eta} - \frac{1}{\eta}$, there are two reference-point consistent strategies at $m = H$ given s^T : $(1, \xi_L^H(\lambda), 1)$ and $(0, 0, \xi_H^H(\lambda))$. Notice that the total utility associated with a strategy that prescribes $(0, 0, \xi_H^H(\lambda))$ after message $m = H$ is 0, while the total utility associated with a strategy that prescribes $(1, \xi_L^H(\lambda), 1)$ after message $m = H$ is $p_H - F - c - \eta(\lambda - 1)p_H(1 - p_H)$. The latter total utility will be higher than 0 if and only if:

$$\lambda \leq 1 + \frac{p_H - F - c}{p_H(1 - p_H)\eta}.$$

Therefore, after message $m = H$, the preferred personal strategy given s^T prescribes $(1, \xi_L^H(\lambda), 1)$ if $\lambda \in (1, \bar{\lambda}(p_H, F, c, \eta))$ and $(0, 0, \xi_H^H(\lambda))$ if $\lambda > \bar{\lambda}(p_H, F, c, \eta)$. If $\lambda = \bar{\lambda}(p_H, F, c, \eta)$, either of these continuation strategies is compatible with the preferred personal strategy given s^T .

6.3 Proof of Corollary 1

Obviously $\underline{\lambda}(p_H, p_L, c, \eta) \leq \bar{\lambda}(p_H, F, c, \eta)$ if and only if:

$$1 + \frac{(c - p_L)(1 + \eta)}{p_H p_L \eta} \leq \max \left\{ 1 + \frac{p_H - F - c}{p_H(1 - p_H)\eta}, \frac{p_H(1 + \eta)}{(F + c)\eta} - \frac{1}{\eta} \right\}$$

Notice that:

$$\begin{aligned} 1 + \frac{(c - p_L)(1 + \eta)}{p_H p_L \eta} &\leq 1 + \frac{p_H - F - c}{p_H(1 - p_H)\eta} \\ &\iff \\ c &\leq \frac{p_L(1 + \eta - p_H\eta - F)}{p_L + (1 + \eta)(1 - p_H)} := c_1(p_L, p_H, F, \eta) \end{aligned}$$

Similarly:

$$1 + \frac{(c - p_L)(1 + \eta)}{p_H p_L \eta} \leq \frac{p_H(1 + \eta)}{(F + c)\eta} - \frac{1}{\eta}$$

$$\iff$$

$$c \leq \frac{p_L - F - p_H p_L + \sqrt{(F + p_L - p_H p_L)^2 + 4p_H^2 p_L}}{2} := c_2(p_L, p_H, F, \eta)$$

Thus $\underline{\lambda}(p_H, p_L, c, \eta) \leq \bar{\lambda}(p_H, F, c, \eta)$ if and only if $c < \bar{c}(p_L, p_H, F, \eta)$, where:

$$\bar{c}(p_L, p_H, F, \eta) := \max \{c_1(p_L, p_H, F, \eta), c_2(p_L, p_H, F, \eta)\}.$$

The second part of the corollary follows immediately from Proposition 5.

6.4 Proof of Proposition 6

Suppose $c < \bar{c}(p_L, p_H, F, \eta)$, $\lambda \in [\underline{\lambda}(p_H, p_L, c, \eta), \bar{\lambda}(p_H, F, c, \eta)]$ and $L < -\frac{G+p_L W}{1-p_L}$. If R thinks that S is following strategy s^T , his behavior is described by Proposition 5. If S knows that the project has high probability of success, she can get $G+p_H W+(1-p_H)L$ by telling the truth or 0 by lying. Assumption 2 implies she will prefer to tell the truth. Now, suppose that the probability of success is low. Then S can get 0 by telling the truth and $G+p_L W+(1-p_L)L$ by lying (this follows from the assumptions we made on λ and on c and from the choice of a tie-breaking rule in which R always play *Stay* whenever indifferent). Our assumption on L implies that S will prefer to announce the truth. Thus a message-dependent and fully-revealing equilibrium exists.

Now suppose that a fully-revealing and message-dependent equilibrium exists. If $\lambda > \bar{\lambda}(p_H, F, c, \eta)$, proposition 5 implies that R's preferred personal strategy given s^T prescribes to play *Out* after any message m contradicting the existence of a message-dependent equilibrium. Thus, we need $\lambda \leq \bar{\lambda}(p_H, F, c, \eta)$ and, as a result, R's preferred personal strategy given s^T prescribes continuation strategy $(1, \xi_L^H(\lambda), 1)$ after message $m = H$. If S knows that the project has low probability of success, she will get payoff 0 by reporting the truth and $(G + \xi_L^H(\lambda) \cdot (p_L W + (1 - p_L)L))$ by lying and overstating the project's quality. Truth-telling requires $G + \xi_L^H(\lambda) \cdot (p_L W + (1 - p_L)L) \leq 0$ and this inequality can be satisfied only if $\lambda \geq \underline{\lambda}(p_H, p_L, c, \eta)$ and $L \leq -\frac{G+p_L W}{(1-p_L)}$. Thus the existence of a message-dependent and fully-revealing equilibrium requires $\lambda \in [\underline{\lambda}(p_H, p_L, c, \eta), \bar{\lambda}(p_H, F, c, \eta)]$ and $L \leq -\frac{G+p_L W}{(1-p_L)}$. Corollary 1 further implies that the condition on λ can be satisfied only if $c < \bar{c}(p_L, p_H, F, \eta)$.

6.5 Proof of Proposition 8

Before proving the actual result, we prove an obvious property of the function $\mu(\cdot)$.

Lemma 1 Let $\mu(\cdot)$ be a function that satisfies properties P1-P4 and P5'. If $px + (1 - p)y < 0$, then $p\mu(x) + (1 - p)\mu(y) < 0$.

Proof. Assume without loss of generality that $y < 0$. If $x < 0$, the result follows immediately from P1 and P2. If $x > 0$, we can have two cases. If $x < |y|$, the result follows from P3. Thus, suppose $x > |y|$. $px + (1 - p)y < 0$ implies $p + p\frac{1}{-y}(x + y) - (1 - p) < 0$. By P5':

$$p + p\frac{1}{-y}(x + y) - (1 - p) > p + p\frac{\mu'(-y)}{\mu(-y)}(x + y) - (1 - p) > p\frac{\mu(x)}{\mu(-y)} - (1 - p)$$

Thus:

$$0 > p\frac{\mu(x)}{\mu(-y)} - (1 - p)$$

which implies

$$0 > p\mu(x) - (1 - p)\mu(-y)$$

and by P3:

$$0 > p\mu(x) + (1 - p)\mu(y)$$

■

We will now prove the statement of the proposition. Throughout the proof S' strategy is given by s^T , we will omit to specify the dependence of strategies and beliefs on it.

Let $m = L$; then $\pi(L) = 0$. Let β^* be a strategy whose continuation strategy after message $m = L$ is given by $\beta^*(L) = (0, x_L(L), x_H(L))$. If R were planning to follow this strategy, his reference utility after message $m = L$ would be given by 0 with probability 1 and his total utility would also be equal to 0. Given this reference utility any strategy whose continuation strategy after message $m = L$ is given by $(e(L), x_L(L), x_H(L))$ would induce a total utility equal to:

$$e(L)x_L(L)p_L - e(L)F - e(L)x_L(L)c + e(L)(1 - x_L(L))\mu(-F) + \\ + e(L)x_L(L)((1 - p_L)\mu(-F - c) + p_L\mu(1 - F - c)).$$

Using Lemma 1 and assumption 1, we can conclude that the previous expression is negative for any value of $e(L)$. It is immediate to see that reference point consistency further requires $x_H(L) = 1$, while the actual level of $x_L(L)$ depends on the comparison:

$$p_L - c + p_L\mu(1 - F - c) + (1 - p_L)\mu(-F - c) \stackrel{\leq}{\geq} \mu(-F)$$

We conclude that there exists a strategy β^* that prescribes *Out* after message $m = L$ and is reference-point consistent at $m = L$ given s^T .

We will now show that this strategy is also the preferred personal strategy. Consider a strategy prescribing $(\hat{e}(L), \hat{x}_L(L), \hat{x}_H(L))$ after message $m = L$. Suppose that $\hat{e}(L) > 0$. In

this case reference point consistency requires the total utility from playing *In* to be at least as big as the one from playing *Out* (if $\hat{e}(L) < 1$ the two total utilities must be equal). In particular, if $\hat{e}(L) > 0$ and $\hat{x}_L(L) = 1$, the total utility would be equal to:

$$p_L - c - F + p_L \hat{e}(L) (1 - p_L) (\mu(1) + \mu(-1)) + (1 - \hat{e}(L)) (p_L \mu(1 - F - c) + (1 - p_L) \mu(-F - c))$$

which is negative by assumption 1, Lemma 1 and P3. Thus a preferred personal strategy cannot prescribe $\hat{e}(L) > 0$ and $\hat{x}_L(L) = 1$. Similarly, if $\hat{e}(L) = 1$ and $\hat{x}_L(L) = 0$, total utility would be equal to $-F$ which is lower than 0; thus the preferred personal strategy cannot prescribe such a continuation strategy after message L . Finally suppose that $\hat{e}(L) > 0$ and $\hat{x}_L \in (0, 1)$. In this case reference point consistency would further require R to be indifferent between playing *Stay* and *Liquidate* after finding out that the probability of success is p_L . Combining this condition with the fact that R is playing *In* with positive probability, the following inequality must hold:

$$(1 - \hat{e}(L)) \mu(-F) + p_L \hat{x}_L(L) \hat{e}(L) (\mu(-1 + c) - \mu(-1 + F + c)) + \\ + (1 - p_L) \hat{x}_L(L) \hat{e}(L) (\mu(c) - \mu(F + c)) \geq F + \hat{e}(L) (1 - \hat{x}_L(L)) \mu(F)$$

By P1 the left-hand side of the inequality is negative, while the right-hand side is positive. We conclude that this strategy is not reference-point consistent. Thus preferred personal strategy must prescribe to play *Out* after message $m = L$.

Now suppose that S announces $m = H$. It is easy to verify that a strategy $(e(\cdot), x_L(\cdot), x_H(\cdot))$ prescribing $e(H) = 1$ can be reference-point consistent only if $x_H(H) = 1$. Furthermore, reference point consistency further requires:

$$x_L(H) = \begin{cases} 1 & \text{if } p_L - c + (1 - p_L) \mu(-1) > \mu(-1 + c) \\ v \in [0, 1] & \text{if } p_L - c + (1 - p_L) \mu(-1) = \mu(-1 + c) \\ 0 & \text{if } p_L - c + (1 - p_L) \mu(-1) < \mu(-1 + c) \end{cases}$$

We conclude that there exists a reference-point consistent strategy at $m = H$ in which $e(H) = 1$ and that the total continuation utility associated with this strategy after message $m = H$ is equal to $1 - F - c > 0$.

Now we will show that, after message $m = H$, the preferred personal strategy must prescribe the continuation strategy we just characterized. Obviously, any strategy $\hat{\beta}(\cdot)$ prescribing $\hat{e}(H) = 0$ would lead to a total utility equal to 0 after message $m = H$ and cannot be part of the preferred personal strategy. Thus we only need to consider continuation strategies that prescribe $\hat{e}(H) \in (0, 1)$. Since $1 - F - c > 0$, reference point consistency requires $\hat{x}_H(H) = 1$.

As a consequence, the total utility associated with this continuation strategy would be equal to $\hat{e}(H) \mu(-1 + F + c) < 0$. We conclude that this cannot be the behavior prescribed by the preferred personal strategy.

References

- Robert J. Aumann. "nash equilibria are not self-endorcing". *Economic Decision Making: Games, Econometric and Optimisation*, pages 201–206, 1990.
- Robert J. Aumann and Sergiu Hart. Long cheap talk. *Econometrica*, 71(6):1619–1660, February 2003.
- Sandeep Baliga and Stephen Morris. Co-ordination, spillovers, and cheap talk. *Journal of Economic Theory*, 105(2):450–468, August 2002.
- Marco Battaglini. Multiple referrals and multidimensional cheap talk. *Econometrica*, 70(4):1379–1401, July 2002.
- Pierpaolo Battigalli and Martin Dufwenberg. Guilt in games. *American Economic Review*, 97(2):170–176, May 2007.
- Pierpaolo Battigalli and Martin Dufwenberg. Dynamic psychological games. *Journal of Economic Theory*, 144(1):1–35, January 2009.
- Pierpaolo Battigalli, Gary Charness, and Martin Dufwenberg. Deception: The role of guilt. *Journal of Economic Behavior & Organization*, Forthcoming, 2013.
- Markus K. Brunnermeier and Jonathan A. Parker. Optimal expectations. *American Economic Review*, 95(4):1092–1118, September 2005.
- Andrew Caplin and John Leahy. Psychological expected utility theory and anticipatory feelings. *The Quarterly Journal of Economics*, 116(1):55–79, February 2001.
- Juan C. Carbajal and Jeffrey Ely. Optimal contracts for loss averse consumers. Mimeo, 2012.
- Gary Charness. Self-serving cheap talk: A test of aumann’s conjecture. *Games and Economic Behavior*, 33(2):177–194, 2000.
- Gary Charness and Martin Dufwenberg. Promises and partnership. *Econometrica*, 74(6):1579–1601, November 2006.
- Gary Charness and Martin Dufwenberg. Participation. *American Economic Review*, 101(4):1211–37, 2011.
- Gary Charness and Martin Dufwenberg. Bare promises: An experiment. *Economics Letters*, 107(2):281–283, 2010.

- Vincent P Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50 (6):1431–51, November 1982.
- David de Meza and David C. Webb. Incentive design under loss aversion. *Journal of the European Economic Association*, 5(1):66–92, 2007. ISSN 1542-4774.
- Wioletta Dziuda. Communication with detectable deceit. Mimeo, January 2011.
- Larry G. Epstein. Living with risk. *Review of Economic Studies*, 75(4):1121–1141, October 2008.
- Joseph Farrell. Talk is cheap. *American Economic Review, Papers and Proceedings*, 85(2): 186–190, 1995.
- Joseph Farrell. Communication, coordination and nash equilibrium. *Economics Letters*, 27 (3):209–214, 1988.
- Joseph Farrell and Robert Gibbons. Cheap talk with two audiences. *American Economic Review*, 79(5):1214–23, December 1989.
- Joseph Farrell and Matthew Rabin. Cheap talk. *Journal of Economic Perspectives*, 10(3): 103–18, Summer 1996.
- Ernst Fehr, Oliver Hart, and Christian Zehnder. Contracts as reference points—experimental evidence. *American Economic Review*, 101(2):493–525, April 2011.
- John Geanakoplos, David Pearce, and Ennio Stacchetti. Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1):60 – 79, 1989. ISSN 0899-8256.
- David Gill and Victoria Prowse. A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102(1):469–503, February 2012.
- David Gill and Rebecca Stone. Fairness and desert in tournaments. *Games and Economic Behavior*, 69(2):346–364, July 2010.
- Maria Goltsman, Johannes Horner, Gregory Pavlov, and Francesco Squintani. Mediation, arbitration and negotiation. *Journal of Economic Theory*, 144(4):1397–1420, July 2009.
- Jerry R. Green and Nancy L. Stokey. A two-person game of information transmission. *Journal of Economic Theory*, 135(1):90–104, July 2007.
- Edoardo Grillo. Reference dependence and politicians’ credibility. Mimeo, 2012.
- Faruk Gul. A theory of disappointment aversion. *Econometrica*, 59(3):pp. 667–686, 1991. ISSN 00129682.

- Oliver Hart and John Moore. Contracts as reference points. *The Quarterly Journal of Economics*, 123(1):1–48, 2008.
- Maxim Ivanov. Communication via a strategic mediator. *Journal of Economic Theory*, 145(2):869–884, March 2010.
- Daniel Kahneman. Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes*, 51(2):296–312, March 1992.
- Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–91, March 1979.
- Daniel Kahneman and Amos Tversky. Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4):1039–61, November 1991.
- Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler. Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy*, 98(6):1325–48, December 1990.
- Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler. The endowment effect, loss aversion, and status quo bias: Anomalies. *Journal of Economic Perspectives*, 5(1):193–206, Winter 1991.
- Navin Kartik. Strategic communication with lying costs. *Review of Economic Studies*, 76(4):1359–1395, October 2009.
- Navin Kartik, Marco Ottaviani, and Francesco Squintani. Credulity, lies, and costly talk. *Journal of Economic Theory*, 134(1):93–116, May 2007.
- Botond Koszegi. Emotional agency. *The Quarterly Journal of Economics*, 121(1):121–156, 2006.
- Botond Koszegi. Utility from anticipation and personal equilibrium. *Economic Theory*, 44(3):415–444, 2010.
- Botond Koszegi and Matthew Rabin. Reference-dependent consumption plans. *American Economic Review*, 99(3):909–36, June 2009.
- Botond Koszegi and Matthew Rabin. A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165, 2006.
- Botond Koszegi and Matthew Rabin. Reference-dependent risk attitudes. *American Economic Review*, 97(4):1047–1073, September 2007.

- Vijay Krishna and John Morgan. The art of conversation: eliciting information from experts through multi-stage communication. *Journal of Economic Theory*, 117(2):147–179, August 2004.
- Vijay Krishna and John Morgan. Cheap talk. In Steven N. Durlauf and Lawrence E. Blume, editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke, 2008.
- Juan S. Lleras. Asymmetric gain-loss preferences: Endogenous determination of beliefs and reference effects. Mimeo, 2012.
- George Loewenstein. Anticipation and the valuation of delayed consumption. *Economic Journal*, 97(387):666–84, September 1987.
- George Loewenstein and Drazen Prelec. Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics*, 107(2):pp. 573–597, 1992. ISSN 00335533.
- Efe Ok, Pietro Ortoleva, and Gil Riella. Revealed (p)reference theory. Mimeo, New York University, 2012.
- Marco Ottaviani and Peter N. Sorensen. Reputational cheap talk. *The RAND Journal of Economics*, 37(1):155–175, 2006. ISSN 1756-2171.
- Matthew Rabin. A model of pre-game communication. *Journal of Economic Theory*, 63(2): 370–391, August 1994.
- Antonio Rosato. Selling substitute goods to loss-averse consumers: Limited availability, bargains and rip-offs. Mimeo, 2012.
- Todd Sarver. Optimal reference points and anticipation. Mimeo, 2012.
- Daniel J. Seidmann and Eyal Winter. Strategic information transmission with verifiable messages. *Econometrica*, 65(1):pp. 163–169, 1997. ISSN 00129682.
- Jonathan Shalev. Loss aversion equilibrium. *International Journal of Game Theory*, 29(2): 269–287, 2000.
- Robert Sugden. Reference-dependent subjective expected utility. *Journal of Economic Theory*, 111(2):172–191, August 2003.