Reference Dependence, Risky Projects and Credible Information Transmission*

Edoardo Grillo†
Princeton University

This version: January 12th 2012.

Job Market Paper

Abstract

We consider the interaction between an informed agent (A) who can make announcements concerning the information she has and an uninformed agent (B) who has to decide whether to start a project or not. We study the role that reference dependence and loss aversion may play in determining the credibility of A's announcement and we show that they may give rise to credible information transmission. This happens because in our model inaccurate information have two effects: they lead B to choose the action A prefers in the short run, but they also generate unrealistic expectations that, through the effect on reference point, may induce B to take long run actions that hurt A. This phenomenon is not possible in a model where the uninformed agent is a standard expected utility maximizer. We further investigate the role that reference dependence can play in settings in which A can write verifiable contractual clauses to reinforce the credibility of her announcements. We show that contractual clauses may lead to credible information transmission, but that this involves monetary transfers that may vary non-monotonically with the degree of loss aversion.

1 Introduction

The transmission of information between an informed Sender and an uninformed Receiver has been the focus of an extensive research agenda in economics; in the most standard model, an informed party (A) has some information that another agent (B) would like to know in order

*I am indebted to my adviser, Stephen Morris, for his advice and encourgement throughout the project. I am also grateful to Roland Bénabou and Wolfgang Pesendorfer for their sugesstions and comments. I also thank all the participants in Microeconomic Theory at Princeton and expecially Dilip Abreu, Sylvain Chassang, Faruk Gul, Juan Ortner and Takuo Sugaya. I take full responsibility for all remaining errors.

†Department of Economics, Fisher Hall, Princeton University, Princeton NJ, 08544-1021 (email: egrillo@princeton.edu)

to take a decision that affects the utility of both.[1] Crawford and Sobel [9] show that, in a one-shot interaction in which lying is costless, information transmission is possible only if the interests of the two agents are sufficiently aligned.

Nevertheless, credible communication appears to be possible also in contexts where the conflict of interests among agents is rather strong: a teamleader does not always lie to the other team members concerning the success probability of a project, even if overstating this probability may increase the effort exerted and advantage him; prospective employees are often provided reasonably accurate descriptions of working conditions, even though some of these details may lead them to reject job offers; although parents may prefer their child to be involved in some social activity (learning to play an instrument, playing a sport and so on), they may provide honest feedback about his's ability to succeed even though this may discourage him.[2]

In environments like the ones described above, information transmission is often justified by the repeated nature of the interaction: the short term gain A could experience by lying is overcome by the long-lasting loss in credibility that could undermine her future utility by preventing any possibility to affect B's behavior in the future. Although dynamic incentives certainly play a key role in determining the credibility of communication, they often involve punishments based on the foregoing of mutually beneficial improvements in subsequent interactions. In particular, even if B is always free to ignore the message sent by A, these punishments may require B to do so also when logic would suggest that A is being sincere.

In this paper we study a different, complementary channel through which credibility can be attained; this channel hinges on the interaction between *reference dependence* and *loss aversion* a là Kösegi and Rabin. Our starting point is the observation that if the announcements made by A modify B's belief, they also modify his future prospects and the utility that he expects to get. Based on an extensive theoretical and experimental literature, we will assume that B's utility depends not only on his material utility, but also on the comparison of this utility with a reference point: any positive (respectively, negative) deviation from this reference point will be associated with a psychological gain (respectively, loss); furthermore we will assume that negative deviations from the reference point hurts the agent more than same-size positive deviations (loss aversion). Thus, in our model A's communication effort will affect B's behavior through two different channels: it will have the standard effect of modifying the probability weight assigned by B to different states, but it will also affect B's beliefs about future prospects and, consequently, his reference point.

In particular, in our model A has some information concerning the quality of a project and B has to decide whether to join the project or not. If B decides to participate in the project, he

[1] In the remaining of the paper, we will use pronoun she to refer to the informed party A and he for the other agent. In line with the literature on strategic communication we will also refer to agent A as to the Sender and to agent B as to the Receiver.

[2] Even if we take the view that the Receiver does not usually trust Sender's announcement, the effort exerted by the Sender party in conveying certain messages suggests that this is not always the case.

learns the true quality of the project and can decide whether to keep working on it, witnessing its success or failure, or to liquidate it. This model is presented in Section 2. To convey the main intuition of the paper, in Section 3 we focus on a case in which: (i) if agents were to care about material utility only, no information transmission would be possible and a positive potential surplus would be wasted, and (ii) the project can be of two types: a high quality project that succeeds with probability 1 and a low quality project that succeeds only with probability $p_L < 1$. Under these assumptions, we show that the introduction of reference-dependent utility and loss aversion in agent B's preferences may lead to the existence of a fully informative equilibrium in which A truthfully reveals the state and B updates his beliefs accordingly. In addition to this fully informative equilibrium, we characterize another, uninformative equilibrium, in which B ignores any message sent by A and does not provide her any incentive to send a particular message.[3]

The mechanism through which reference dependence and loss aversion can induce truthful information transmission can be summarized as follows. Suppose B has reference dependent utility; then A's announcement, if credible, affects his beliefs concerning the quality of the project and, consequently, his future prospects, namely his reference utility. Thus, if B were to find out that A lied (by claiming that the project is a high quality one while it is not), he may decide to keep working in order to avoid the psychological loss associated with giving up, even if this behavior is suboptimal from an ex-ante (pre-communication) point of view.[4] If the decision to keep working on bad quality projects is sufficiently harmful for A, she will prefer not to lie and this would lead to credible information transmission.

We want to stress that, in the mechanism we are proposing, B is taking a suboptimal action from an ex-ante perspective (namely, keeping working on bad quality projects), but once communication has taken place and reference points have been determined accordingly, the behavior of B will be the utility maximizing. Indeed, the effect of communication on B's future utility is what makes the B's "punishment strategy" endogenously credible. Intuitively, whenever B finds out that A lied about the quality of the project, he faces a trade-off: he can take the material utility maximizing action accepting to incur the psychological loss associated with a negative deviation from his reference point (which was based on the false announcement made by A), or he can take an action that is suboptimal from the material point of view but can, with some probability, reduce the psychological loss. If the degree of loss aversion is sufficiently large, B will prefer to follow the second strategy and this will endogenously discipline A to tell the truth. What drives the type of behavior described above is the change in the attitude toward risky lotteries experienced by B through loss aversion;

[3] The existence of an uninformative equilibrium together with informative ones is standard in the literature on strategic communication and follows from the freedom that B has in ignoring A's message.

[4] In our model when B decides whether to keep working on the project or to give up, he knows the actual probability of success. Therefore, the behavior described above is not determined by an incorrect evaluation of the probability of success. For more details on this, see Section 4.1.

faced with future prospects that are worse than expected, B will be more willing to take a risky choice that may decrease the probability of incurring in a loss. Since B's expectation are determined by A's speech, if A dislikes this risky choice, this mechanism will provide credibility to her announcements.

In Section 4, we relax some of the assumptions of the baseline model to study the robustness of the mechanism described above. In particular, we consider extensions of the model in which we allow for randomness in the probability with which B finds out the actual state and in the profitability of good quality projects. In all these cases, reference dependence can still induce credible information transmission, even though some additional restriction on parameters is necessary. Finally, in this paper we focus on a model in which reference dependence and loss aversion in the Receiver's utility facilitates communication, but Section 4 discusses both the possibility that this type of utility can prevent communication instead of facilitating it and the case in which also the Sender has reference-dependent attitudes.

The kind of communication described in our baseline model is not verifiable and agent A cannot be held responsible for its veridicity; this may represent a feature of the chosen communication channel (oral communication, non-binding written agreements) or of the actual content of communication (some information, although observable by the agents involved, may be hard to verify in a court of justice: think, for example, of the "true" probability of success of a project). In order to understand the role that reference dependence and loss aversion can play in more complex settings, Section 5 allows A to back her announcements with verifiable and enforceable monetary transfers; in particular, we show that these enforceable transfers can induce credible information transmission on a larger set of parameters and that they may vary non-monotonically with the degree of loss aversion: whereas they may be high for low and high values of loss aversion, they become 0 for intermediate degrees.

In the remaining of the introduction, we review the relevant literature.

1.1 Related Literature

In this paper we analyze the issue of credible information transmission in a setting where the receiver has reference-dependent preferences. Thus our paper is related to the literature on strategic communication and to one on reference dependence and loss aversion.

Strategic information transmission has been the focus of a massive literature.[5] The seminal work of Crawford and Sobel [9] has shown that in a static setting,[6] as long as lying is costless, information transmission can be attained only if the interests of the informed and preferences

[5] See, for example, Crawford and Sobel [9], Farrell and Gibbons [14], Farrell [13], Battaglini [2], Aumann and Hart [1], Krishna and Morgan [35], [36] and the references therein. For surveys on the literature, see Farrell and Rabin [15] and Krishna and Morgan [37].

[6] For a different approach see Green and Stokey [19].

of the parties are sufficiently aligned.[7] In the model of this paper, information transmission is made difficult by the fundamental conflict of interest between agents in the low state.

In order to overcome the impossibility of transmitting information in settings were the conflict of interest is high, the literature often uses the multiperiod structure of many interactions. Indeed, in a repeated game setting, although lying may lead to one-shot gains, it may also undermine Sender's long-term credibility and, consequently, her future gains.[8] Following a different approach and keeping the static nature of the game fixed, Aumann and Hart [1] analyze the set of payoff attainable with arbitrary rounds of communication. In our paper, despite the dynamic structure of the interaction, lies do not affect A's future credibility and, with standard expected utility, B would not be able to commit to any punishment strategy against A; furthermore, the addition of multiple rounds of communication would not overcome the difficulties associated to information transmission.[9]

Alternatively, Goltsman et al. [18] and Ivanov [23] have shown that the conflict of interests between the Sender and the Receiver can be overcome by introducing a mediator whose role is to weaken the link between the announcements sent by the informed party and the action taken by the uninformed agent.[10] Since, in the baseline model of this paper, the uninformed agent perfectly learns the state, the introduction of a mediator would not help in establishing the credibility of the Sender.

Ottaviani and Squintani [40] establish the credibility of information transmission by introducing naive agents who interpret literally the announcements sent by the Sender without taking into account her incentive to distort their actions. Whereas the uninformed agent of our model is subject to a behavioral bias, the mechanism through which this bias discipline the informed agent is deeply different. In our model, agent B is fully aware that A has an incentive in distorting his action, but the effect of A's words on B's reference point will lead to a change on B's attitude toward the continuation of the project.

Dziuda [11] analyzes a model in which lies are detectable with some exogenous probability and she looks at the effect of this probability on information transmission. Similarly to her model, the possibility of detecting a lie play an important role in our model, but the actual mechanism through which this happens is different.[11]

The idea that people evaluate the consequences of their actions with respect to a reference point and that they exhibit loss aversion have been formalized by Kahneman and Tversky [27].

[7] To understand the implications of relaxing the assumption of costless lying, see Kartik [28] and Kartik et al. [29].

[8] The intuition behind this result is particularly disturbing in settings where the equilibrium construction would require the uninformed agent to ignore the announcements (reverting to a babbling equilibrium) even when the logic would suggest that the informed agent would be willing to tell the truth.

[9] Abstracting from its dynamic structure, our game is actually similar to Example 2.2 in Aumann and Hart [1].

[10] Goltsman et al. [18] deal with an unbiased mediator, while Ivanov [23] considers the case of a biased one.

[11] See Section 4.1 for further details.

Although several authors have accepted the assumption that agents have reference-dependent preferences, they disagreed on the actual specification of reference points. Whereas some authors have taken a backward-looking approach assuming that the reference point is based on an agent's status quo,[12] other scholars have assumed that expectations and future prospects are key in determining the utility an agent feels entitled to.[13] This latter approach raises the additional issue about how to close the loop between agents' optimality and reference point's formation. In a series of papers, Köszegi and Rabin [32], [33], [34] close this loop by assuming rational expectations: in equilibrium, reference points are determined by agents' beliefs under the assumption that they behave optimally, and the optimality of a strategy is assessed taking into account its effect on the formation of reference point.[14] In this paper, we follow Köszegi and Rabin's insight, but we further look at the equilibrium effect that A's announcements may have on B's reference point. In doing this, we adapt to our setting the different solution concepts proposed by Köszegi and Rabin [33], [34].[15]

Köszegi [30] looks at the effect of communication in a model in which one of the agents have anticipatory utilities and the interest of the two parties are perfectly aligned.[16] In the present paper, as well as in Grillo [20], we focus on the conflict of interests between the two agents and we look at its implications on credible information transmission. Furthermore, the assumption of reference-dependent preferences allows us to describe the specific behavior of B following a lie.

In our paper, beliefs concerning the state of nature affect the preferences of agents even after controlling for the actual quality of the project; thus, the utility of agents depends on beliefs determined at previous nodes in the game. In this respect, our paper is related to the literature on psychological games pioneered by Geanakoplos et al. [17] and extended to dynamic settings by Battigalli and Dufwenberg [4].[17]

Hart and Moore [22] and Fehr et al. [16] study the effect of reference dependence on the choice between formal and informal contracting in a theoretical and experimental setting. Although in our model we do not specifically address this issue, we study the circumstances under

[12] See Kahneman and Tversky [44], Kahenman [24], Kahneman et al. [26], [25], Sugden [43] and the references therein.

[13] See Shalev [42] and the references cited below. In a similar way, Gul [21] 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.

[14] See Section 2.1 for further details.

[15] See also Kösegi [31]. In particular, the "surprise" situations described in Kösegi and Rabin [33], [34] correspond in our model to cases in which the Receiver finds out that the Sender lied when he was not expecting her to do so (these are situations in which the Receiver updates his beliefs so to assign probability 0 to a state of nature, but he then finds out that this state of nature is the relevant one). Furthermore, in the definition of Köszegi and Rabin's Preferred Personal Equilibrium, the probability of different decision sets in exogenously given, while in our model is determined in equilibrium also by the communication of the Sender.

[16] On anticipatory utilities see Loewenstein [38] and Loewenstein and Prelec [39]. For an axiomatic treatment of anticipatory utilities see Caplin and Leahy [6] and Epstein [12].

[17] See also Rabin [41] and Battigalli and Dufwenberg [3].

which the endogenous formation of the reference point can provide credibility to informal announcements and we further describe the role that enforceable monetary transfers can play in this setting.

De Meza and Webb [10] investigate the effect of loss aversion in a principal-agent model under different assumptions on the formation of the reference point. For each of these assumptions, they characterize the optimal compensation scheme and they show that it may not be strictly increasing in performance. Although our focus is on credible communication, Section 5 shows that monetary transfers interact with loss aversion in a nontrivial way and this may lead to a non-monotonic relationship between monetary transfers and the coefficient of loss aversion.

Finally, Charness and Dufwenberg [7] and [8] provide experimental evidence showing that communication may affect the attitude of agent toward participation in risky projects. Although the model we analyze share some common features with theirs, the channel through which communication affects the behavior is different: whereas Dufwenberg and Charness look at the role of guilt, we consider an intention-free setting in which the behavior of the Receiver is affected by the change in the reference point.

2 The Model

Two agents, A and B, are involved in a joint project. The probability of success of the project depends on its quality and we assume that there are two types of projects: high quality projects (denoted with θ_H) and low quality ones (denoted with θ_L). We further assume that A knows the true quality of the project, while B does not and assigns probability $\frac{1}{2}$ to each possibility.

The timing of the model is as follows:

- in period $t = -1$, nature chooses the quality of the project
- in period $t = 0$, A can send a message concerning the quality of the project.
- in period $t = 1$, upon listening to A's announcement, B decides whether to enter in the partnership (action In) or stay (action Out). In the former case B incurs a cost of c_1, A gets a payoff of G and the game moves to period $t = 2$. In the latter case the game is over and both agents get an outside utility that we normalize to 0.
- in period $t = 2$, B learns the true quality of the project and decides whether to keep working on the project (action *Stay*) incurring an additional cost of c_2 or to liquidate the project ending the game (action *Liquidate*).
- in period $t = 3$, a random variable determines whether the project succeeds (outcome s) or fails (outcome f).

In the baseline model, we will assume that the probability of success is given by $p_H = 1$ in state θ_H and by $p_L < 1$ in state θ_L, but Section 4.2.1 will relax this assumption allowing for uncertainty in the success probability of good quality projects. We further assume that agents do not discount the future and that their initial utilities are equal to 0. Figure 1 summarizes the structure of the game.

The material utility of B is the sum of two components: the outcome-related payoff associated with the success or failure of the project and the cost associated with the effort he exerts. Although we label these two components as outcome-related and effort, we can interpret the effort component as earlier payoff and the outcome-related component as later payoff.[18] To simplify notation, we will denote with C the total cost B incurs if he joins the project and keeps working on it, that is $C = c_1 + c_2$. The outcome-related payoff experienced by B is equal 1 if the project succeeds and to 0 if the project fails. To make the transmission of information relevant, we introduce the following assumption concerning payoffs:

Assumption 1 *(i)* $p_L < c_2 < p_H$, *(ii)* $C < p_H$, *(iii)* $\frac{p_H}{2} < c_1 + \frac{c_2}{2}$

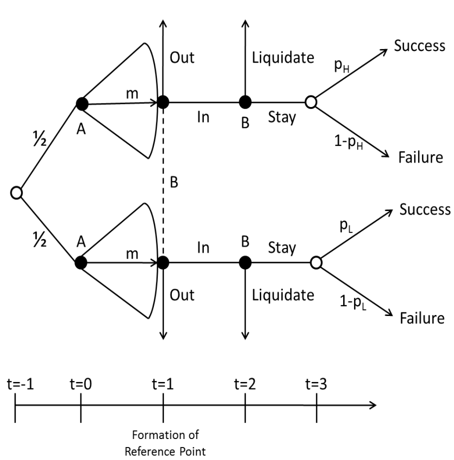


Figure 1: The Structure of the Game

To understand these assumptions, suppose that B is a risk neutral agent who experiences a linear disutility from exerting effort. Assumption 1(i) states that, conditional on having to choose between liquidation and continued engagement in the project, B would abandon low quality projects and keep working on good quality ones. Assumption 1(ii)-(iii) states that the total cost is sufficiently low to guarantee participation if B is certain that the project is of good quality (ii), but also that this cost is sufficiently high to prevent participation when B

[18] This termporal interpretation could require to distinguish between c_1 and c_2 as well. None of the results would be affected by this additional distinction.

has no information about its true quality (iii). Therefore, if A does not convey any credible information concerning the quality of the project, B will prefer to choose action Out from the beginning. Observe that the decision to keep working in the project is both costly and risky: with some probability B will get a positive outcome-related payoff, but with complementary probability, he would get nothing and would waste the additional effort exerted. An important feature of this model is that $p_L > 0$, so that low quality projects succeed with some positive probability.

We assume that A experiences a payoff $G > 0$ whenever B participates to the project and that he gets an additional payoff of $S > 0$ if the project succeeds and of $L < 0$ if the project fails.[19]

In the model we are describing A is active in period 0 only and in that period he makes an announcement concerning the quality of the project. Let M be the finite set of messages available to agent A; then the behavior of A can be represented by a function $t : \{\theta_L, \theta_H\} \to M$.[20] To formally describe the behavior of B, we need to introduce some further notation concerning the structure of the game. Agent B is active at two different information sets: (i) upon listening to A's announcement, he has to decide whether to participate in the project or not, and (ii) upon learning the true state he has to decide whether to keep working on the project or to liquidate it. If we denote each history with the profile of actions that leads to it,[21] the first class of information sets will be denoted with $\mathcal{I}_{\mathcal{M}}$, where

$$\mathcal{I}_{\mathcal{M}} = \{\{(\theta_L, m), (\theta_H, m)\} : m \in M\}.$$

The second class of information sets is denoted with $\mathcal{I}_{\mathcal{M},\Theta}$ and is defined by:

$$\mathcal{I}_{\mathcal{M},\Theta} = \{(\theta_i, m, In) : m \in M,\ i \in \{L, H\}\}$$

To simplify the notation, we will denote information set $\{(\theta_L, m), (\theta_H, m)\} \in \mathcal{I}_{\mathcal{M}}$ with m and information set $(\theta_i, m, In) \in \mathcal{I}_{\mathcal{M},\Theta}$ with (m, θ_i). Finally for any $m \in M$, we will denote with $\mathcal{I}_{\mathcal{M},\Theta}(m)$ the set of information sets in $\mathcal{I}_{\mathcal{M},\Theta}$ compatible with message m (or equivalently to the fact that agent B is at information set $m \in \mathcal{I}_{\mathcal{M}}$); thus:

$$\mathcal{I}_{\mathcal{M},\Theta}(\bar{m}) = \{(\theta_i, m, In) : m \in \{\bar{m}\}\ i \in \{L, H\}\}$$

Then, the strategy of B can be represented by a behavioral strategy (α, β), where $\alpha : \mathcal{I}_{\mathcal{M}} \to$

[19] Equivalently, we could assume that the payoff from the project are s and l depending on the failure or success of the project and that c_A represents a cost A incurs if the project is not liquidate. Clearly, we can redifine $S = s - c_A$ and $L = l - c_A$. In this case, the assumption $L < 0$ implies that the cost associated with the return from the project is low compared to its cost if the project fails.

[20] Given a finite set X, we denote with $\Delta(X)$, the set of probability measures over X.

[21] Thus, for example, (θ_i, m, Out) will correspond to the history in which Nature chose θ_i, A sent message m and B played Out.

$[0,1]$ and $\alpha(m)$ is the probability with which B chooses In at information set m, while $\beta : \mathcal{I}_{\mathcal{M},\Theta} \rightarrow [0,1]$ and $\beta(m,\theta_i)$ represents the probability with which the agent chooses $Stay$ at information set (θ_i, m, In). Clearly, the analysis of B's behavior requires to specify his belief concerning the state of nature at each of the information set in which he is active. Suppose B believes has conjecture $\tilde{t} \in \Delta\left(M^{\{\theta_L,\theta_H\}}\right)$ on the strategy of agent A; then $\pi\left(m;\tilde{t}\right)$ will be the probability that B assigns to state θ_H at information set m and will be determined by Bayes rule as follows:

$$\pi\left(m;\tilde{t}\right) = \frac{\sum\limits_{t:t(\theta_H)=m} \tilde{t}[t]}{\sum\limits_{\theta\in\{\theta_L,\theta_M\}} \sum\limits_{t:t(\theta)=m} \tilde{t}[t]} \tag{1}$$

Given the assumption that B learns the state after entering in the project, we can denote with $\pi(m,\theta_i)$ the probability that B assigns to state θ_H at information set (m,θ_i). We assume that:[22]

$$\begin{aligned} \pi(m,\theta_H) &= 1 \ \forall m, & (2) \\ \pi(m,\theta_L) &= 0 \ \forall m. & (3) \end{aligned}$$

We will refer to $\left(\pi\left(.;\tilde{t}\right), \pi(.,\theta_H), \pi(.,\theta_H)\right)$ as to the belief system induced by $\tilde{t}$ and we will denote it with $\pi\left(\tilde{t}\right)$.

We will now analyze the model assuming that A is a standard expected utility maximizer, but considering two different types of utility for agent B: standard expected utility and reference-dependent utility. Before moving to the formal analysis of the model, we provide a short discussion of reference dependent utility in the context of our model.

2.1 Reference-Dependent Preferences

In this paper we use the utility function introduced by Köszegi and Rabin [32], [33], [31] 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:

$$v(a \mid r) = u(a) + \mu(u(a) - u(r)), \tag{4}$$

where:

$$\mu(x) = \eta \cdot \max\{0,x\} + \eta\lambda \min\{0,x\} \ \ \forall x \in \mathbb{R} \tag{5}$$

[22] Observe that 1 and 2 imply that even if the agent were assigning probability 1 to state θ_i after A's announcement, he will change his mind if the hard evidence he receives in period 2 states otherwise.

with $\eta \in (0,1)$, and $\lambda > 1$. Thus, the utility of an agent with reference-dependent preferences is represented by a function $v(. \mid .) : Z \times Z \rightarrow \mathbb{R}$, where the first argument, a, is the actual outcome and the second, r, is the reference outcome. In particular, utility function $v(a \mid r)$ is the sum of two components: (i) the material or consumption utility represented by utility index $u(.)$ and (ii) the gain/loss component represented by function $\mu : \mathbb{R} \rightarrow \mathbb{R}$. The gain/loss component captures the idea that agents evaluate outcome a with respect to reference point r; to be more precise, whenever the utility associated with outcome a, $u(a)$, exceeds (respectively, falls short of) the reference utility $u(r)$, the agent experiences a psychological gain (respectively, a psychological loss); parameter η measures the relative importance of psychological gains with respect to material utilities. Furthermore, (4) and (5) capture the idea that agents are loss averse, that is, they suffer from losses more than how they benefit from gains of the same size.

(4) and (5) can be extended to account for random outcomes $\widetilde{a} \in \Delta(Z)$ given a fixed reference point $r \in Z$ in the following way:

$$\forall \tilde{a} \in \Delta(Z), \ v(\tilde{a} \mid r) = \sum_{a \in Z} v(a \mid r) \tilde{a}(a),$$

and for random outcomes and reference points as follows:

$$\forall \tilde{a}, \tilde{r} \in \Delta(Z), \ v(\tilde{a} \mid \tilde{r}) = \sum_{r \in Z} \sum_{a \in Z} v(a \mid r) \, d\tilde{a}(a) \, d\tilde{r}(r),$$

Finally we can extend the definition of reference dependent utilities to multidimensional outcome spaces: let $Z = Z_1 \times Z_2 \times ... \times Z_n$ be the set of n-dimensional outcomes; an element of Z is a n-dimensional vector $(a_1, a_2, ..., a_n) \in Z$. For each pair $\tilde{a}, \tilde{r} \in \Delta(Z)$, the reference dependent utility is given by:

$$\forall \tilde{a}, \tilde{r} \in \Delta(Z), \ v(\tilde{a} \mid \tilde{r}) = \sum_{r \in Z} \sum_{a \in Z} \left(\sum_{i=1}^{n} v_i(a_i \mid r_i) \right) d\tilde{a}(a) \, d\tilde{r}(r) \tag{6}$$

where for each $i \in \{1, ..., n\}$, $v_i(a_i \mid r_i)$ is defined as in (4). Observe that (6) implies that B has separable and additive utilities over the different dimensions.

In the previous definitions, the reference point has been taken as exogenous. One of the main contribution of Köszegi and Rabin [32], [33], [34] is to endogenize the reference point through equilibrium analysis. To understand how this can be accomplished, consider a static decision problem in which a decision maker has to choose an element from a finite set of options D. Let $\zeta : D \rightarrow Z$ be the *outcome function* mapping each decision into a final outcome and assume that, whenever the decision maker chooses outcome d, he foresees inducing outcome $\zeta(d)$. Thus, following, Köszegi and Rabin [32], [33], [34] we can introduce two different definitions of optimality.[23] The first evaluates the optimality of an action without taking into account the effect of possible deviations on the reference point: a choice $d \in D$ is optimal if for any other $d' \in D$

$$v\left(\zeta\left(d\right) \mid \zeta\left(d\right)\right) \geq v\left(\zeta\left(d'\right) \mid \zeta\left(d\right)\right) \tag{7}$$

The second definition of optimality incorporates the effect of deviations on the reference point and selects among all decisions satisfying (7) the one that maximizes total utility. Formally, action $d \in D$ is optimal if it satisfies (7) and for any other $d' \in D$ satisfying (7),

$$v\left(\zeta\left(d\right) \mid \zeta\left(d\right)\right) \geq v\left(\zeta\left(d'\right) \mid \zeta\left(d'\right)\right). \tag{8}$$

Given the dynamic structure of the model we analyze in this paper, the formation of reference utility and the definition of optimality requires some additional discussion. Suppose that B believes that A is following communication strategy t; then, we can determine a belief $\pi\left(m;t\right)$ according to 1 for each possible information set m. If B were to play behavioral strategy $\left(\alpha,\beta\right)$, his material utility at information set m would be a random lottery:

$$\tilde{u}\left(\alpha,\beta;m,t\right) = \left(\tilde{u}_1\left(\alpha,\beta;m,t\right),\tilde{u}_2\left(\alpha,\beta;m,t\right)\right)$$

where

$$\tilde{u}_1\left(\alpha,\beta;m,t\right)[x] = \begin{cases} \pi\left(m,t\right)\alpha\left(m\right)\beta\left(m,\theta_H\right) + \left(1-\pi\left(m,t\right)\right)\alpha\left(m\right)\beta\left(m,\theta_L\right)p_L & \text{if } x=1 \\ \\ 1-\alpha\left(m\right)\left(\pi\left(m,t\right)\beta\left(m,\theta_H\right) + \left(1-\pi\left(m,t\right)\right)\beta\left(m,\theta_L\right)p_L\right) & \text{if } x=0 \end{cases}$$

is the random lottery in the outcome-related component and:

$$\tilde{u}_2(\alpha,\beta;m,t)[x] = \begin{cases} 1-\alpha(m) & \text{if } x=0 \\ \\ \pi(m;t)\alpha(m)(1-\beta(m,\theta_H)) + (1-\pi(m;t))\alpha(m)(1-\beta(m,\theta_L)) & \text{if } x=-c_1 \\ \\ \pi(m;t)\alpha(m)\beta(m,\theta_H) + (1-\pi(m;t))\alpha(m)\beta(m,\theta_L) & \text{if } x=-C \end{cases}$$

is the random lottery in the cost component.

Similarly the random utility of B when he plays strategy (α,β) at information set (m,θ_i) is given by:

$$\tilde{u}\left(\beta;m,\theta_i\right) = \left(\tilde{u}_1\left(\beta;m,\theta_i\right),\tilde{u}_2\left(\beta;m,\theta_i\right)\right),\ i \in \{L,H\}$$

[23] Conditions that guarantee the existence of these equilibria are provided in Köszegi [31].

where for every $i \in \{L, H\}$:

$$\tilde{u}_1(\beta; m, \theta_i)[x] = \begin{cases} \beta(m, \theta_i) p_i & \text{if } x = 1 \\ \\ 1 - \beta(m, \theta_i) p_i & \text{if } x = 0 \end{cases}$$

and

$$\tilde{u}_2(\beta; m, \theta_i)[x] = \begin{cases} \beta(m, \theta_i) & \text{if } x = -C \\ \\ 1 - \beta(m, \theta_i) & \text{if } x = -c_1 \end{cases}$$

The formation of B's reference point deserves some further comments. If B is playing behavioral strategy (α, β) and has conjecture t about the behavior of agent A, his reference utility will be given by:

$$\tilde{v}(\alpha, \beta; m, t) = (\tilde{v}_1(\alpha, \beta; m, t), \tilde{v}_2(\alpha, \beta; m, t)),$$

where $\tilde{v}(\alpha, \beta; m, t)$ is defined analogously to $\tilde{u}(\alpha, \beta; m, t)$.

Let $\tilde{u}$ and $\tilde{v}$ be two finite lotteries over real number. Then, we simplify notation by defining:

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

and

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

Suppose that B believes A is following behavioral strategy t. Then the total utility at information set m, given conjecture t and reference point $\tilde{v}$ is equal to:

$$v(\alpha, \beta \mid m, t, \tilde{v}) = E\tilde{u}(\alpha, \beta; m, t) + \mu(\tilde{u}(\alpha, \beta; m, t) - \tilde{v})$$

Similarly the total utility at information set (m, θ_i) is given by:

$$v(\beta \mid m, \theta_i, \tilde{v}) = E\tilde{u}(\beta; m, \theta_i) + \mu(\tilde{u}(\beta; m, \theta_i) - \tilde{v})$$

So far, we introduced B's actual utility and reference utility as if they were determined independently. However, in equilibrium, behavioral strategy (α, β) affects the reference utility and the optimality of (α, β) has to be evaluated taking into account the reference utility induced by that strategy. We say that behavioral strategy (α, β) is *dynamic consistent* if this strategy is optimal given the reference utility generated by (α, β) itself. To understand this point, observe that at information set $m \in \mathcal{I}_{\mathcal{M}}$, agent B: (i) modifies his belief about the actual quality of the project, and (ii) formulates a plan about how to behave for any possible information set he may be at in period 1. *Dynamic consistency* requires strategy (α, β) to be optimal given that reference utility is equal to $\tilde{v}(\alpha, \beta; m, t)$.[24] Formally:

Definition 1 *Behavioral strategy (α, β) is dynamically consistent given conjecture t at information set $m \in \mathcal{I}_{\mathcal{M}}$, if $\forall \hat{\alpha} \in [0,1]^M$:*

$$v(\alpha, \beta \mid m, t, \tilde{v}(\alpha, \beta; m, t)) \geq v(\hat{\alpha}, \beta \mid m, t, \tilde{v}(\alpha, \beta; m, t)) \tag{9}$$

and $\forall \theta_i \in \{\theta_L, \theta_H\}$ and $\hat{\beta} \in [0,1]^{M \times \Theta}$:

$$v(\beta \mid m, \theta_i, \tilde{v}(\alpha, \beta; m, t)) \geq v\left(\hat{\beta} \mid m, \theta_i, \tilde{v}(\alpha, \beta; m, t)\right) \tag{10}$$

Behavioral strategy (α, β) is dynamically consistent given t if (α, β) is dynamically consistent given t for every $m \in M$.

Thus, dynamic consistency tests the optimality of strategy (α, β) under the assumption that the reference utility is fixed to the one that strategy (α, β) would induce; indeed possible deviations are evaluated without taking into account the effect of these deviations on the reference utility. In this sense, dynamic consistency captures the same optimality requirement of (7).

Observe that Definition 1 does not rule out the existence of a different dynamic consistent strategy $\left(\hat{\alpha}, \hat{\beta}\right)$ that at some information set m does better than (α, β) once we take into account that the reference utility associated with $\left(\hat{\alpha}, \hat{\beta}\right)$ is given by $\tilde{v}\left(\left(\hat{\alpha}, \hat{\beta}\right); m, t\right)$. This issue is addressed in the following definition:[25]

Definition 2 *Strategy (α^*, β^*) is the optimal dynamic consistent strategy given t if (i) (α^*, β^*) is dynamic consistent and, (ii) for any other dynamic consistent strategy $\left(\hat{\alpha}, \hat{\beta}\right)$:*

$$v(\alpha^*, \beta^* \mid m, t, \tilde{v}(\alpha^*, \beta^*; m, t)) \geq v\left(\hat{\alpha}, \hat{\beta} \mid m, t, \tilde{v}\left(\hat{\alpha}, \hat{\beta}; m, t\right)\right) \tag{11}$$

for every m.

Thus, the definition of an optimal dynamic consistent strategy evaluates deviations taking into account their effect on the reference utility. Therefore, the definition of an optimal dynamic consistent strategy incorporates the optimality criterion contained in (8).

We are now ready to introduce the equilibrium definition that we will use in this paper.

[24] Kösegi [31] and Kösegi and Rabin [32], [33] and [34] refers to dynamic consistent strategies as to *personal equilibria*; we decided to use a different name to stress the dynamic nature of these requirements and to highlight the distinction between a consistent strategy of a player and a game theoretic equilibrium.

[25] In the terminology of Kösegi [31] and Kösegi and Rabin [32], [33] and [34], an optimal dynamic consistent strategy is called *preferred personal equilibrium.*

Definition 3 *A profile of behavioral strategies* $(t^*, (\alpha^*, \beta^*))$ *and a belief system* $\pi(t^*) = (\pi(.;t^*), \pi(.,.))$ *is an equilibrium if:*
(i) (α^*, β^*) *is the optimal dynamic consistent strategy for agent B given belief system* $\pi(t^*)$.
(ii) t^* *maximizes A's utility given* (α^*, β^*).
(iii) $\pi(t^*)$ *is determined according to 1, 2 and 3 given* t^*.

Thus, our solution concept requires the agent to choose the optimal dynamic consistent strategy; to put it differently, whenever the agent has more than one dynamic consistent strategy, our solution concept will require him to select the one associated with the highest utility.[26] In this paper we will deal with two class of equilibria: fully informative ones and uninformative ones. We provide a definition of these two equilibria below:

Definition 4 *Let* $(t^*, (\alpha^*, \beta^*))$ *be an equilibrium. Then:*
- $(t^*, (\alpha^*, \beta^*))$ *and* $\pi(t^*)$ *is fully informative if* $\{m : t(\theta_H) = m \text{ for some } t \text{ with } t^*[t] > 0\} \cap \{m : t(\theta_L) = m \text{ for some } t \text{ with } t^*[t] > 0\} = \emptyset$.
- $(t^*, (\alpha^*, \beta^*))$ *and* $\pi(t^*)$ *is uninformative if for every* m $\sum_{t:t(\theta_H)=m} t^*[t] = \sum_{t:t(\theta_L)=m} t^*[t]$.

Thus, in a fully informative equilibria, B has (correct) degenerate beliefs after listening to A's announcement, while in an uninformative equilibrium A's announcements have no informational content and B does not update his beliefs upon listening to them.
Finally, observe that in the particular case in which $\eta = 0$ (no psychological loss or gain), the definition of dynamic consistent strategy and optimal dynamic consistent strategy coincide and are equivalent to the standard optimality requirement of dynamic games. We summarize this observation in the following Remark:

Remark 1 *If* $\eta = 0$, *conditions (9) is equivalent to (11). Furthermore (9) and (10) the definition of equilibrium coincides with that of a Perfect Bayesian Equilibrium.*

We conclude this section by pointing out that the announcement of A, *if credible*, will modify the belief concerning the state of nature, $\pi(m;t)$, and will induce a change in B's behavior through two different channels: (i) it will modify the expected utility associated with different actions by changing the probability weight associated with states of nature, and (ii) it will modify the reference utility of B and affecting the way in which he evaluates the optimality of different strategies. Note that, if at some information set $m \in \mathcal{I}_M$, agent B were to assign probability 0 to some $(m, \theta_i) \in \mathcal{I}_{M,\Theta}(m)$, the behavior prescribed by β at (m, θ_i) would not play a role in determining the reference utility. This will happen either if $\alpha = 0$ (in which case

[26] Although we model the choice of the reference point as a conscious act of the agent, we could equivalently interpret it as a totally unconscious process. Furthermore, we can relax the assumption that the agent always selects the "optimal" strategy with the assumption that the optimal strategy is chosen only with some positive probability. For a model on the choice of beliefs, see Brunnermeier and Parker [5].

β would be irrelevant at any information set) or if $\pi(m;t) \in \{0,1\}$ (if $\pi(m,t) = 0$, $\beta(m,\theta_H)$ would not affect the reference utility, while if $\pi(m,t) = 1$, $\beta(m,\theta_L)$ would be irrelevant).

3 Reference Dependence and Truth-telling

In the baseline model that we analyze in this section, we make the following assumptions:

Assumption 2 *High quality projects always succeed:* $p_H = 1$.

Assumption 2 states that high quality projects always succeed. Although this assumption is not necessary, it will help highlighting the mechanism behind our result.[27]

Assumptions 1 and 2 are made to focus our attention on the case in which the inability of A to convey information concerning the quality of the project leads to inefficiently low participation in state θ_H. Finally, if $p_L > C$ (respectively, $C > 1$), B would (respectively, would not) participate in the project regardless of its actual quality. In both cases, the information sent by A would not affect the behavior of B.

Although stylized, the previous assumptions fit several type of interactions, some of which are described below.

Example 1 *An entrepreneur (A) has a project, but needs to raise money from another agent (B) to finance or implement it. In this case,* c_1 *and* c_2 *can represent both the monetary disbursements or the actual effort exerted by B in financing the project. Agent A knows the true quality of the project, while B does not. In this context* G *can represent the positive amount of money A can divert to her own account if B finance the project or a direct gain coming from relaxing the liquidity constraint. The loss* $-L$ *experienced by A if the project fails can be interpreted as some type of effort or as some bankruptcy cost associated with the failure of the project.*

Example 2 *A prospective employee (B) is deciding whether to accept a job offer or not; the offer entails some fixed salary equal to* L *and a bonus equal to* R*, where* $R = 1 - L$ *if the project is successful.* c_1 *and* c_2 *represents the costs associated with the job (depending on the context, these may represent relocation cost, effort cost, cost opportunity of outside options). The potential employer (A) knows whether the working conditions are good or not. Good working conditions enable B to succeed in his task with higher probability. If B accepts the job he learns the working conditions and can decide whether to keep working or to resign. A experiences a benefit equal to* G *if B accepts the job (we can think of A as being able to steal some of the B's know-how by hiring him) and she gets an additional payoff equal to* $X = S + L + R > 0$ *if the project succeeds and no additional payoff if the project fails.*

[27] We will relax it in Section 4.2.1.

Example 3 *A child (B) has to decide whether to initiate a new hobby or activity (playing a new instrument, enrolling in a sport team). One of his parents (A), knows the talent of the child and experiences a positive utility (G) if the child is engaged in the activity (she may assign positive utility to the child being involved in socializing activities), and she experiences a positive (respectively, negative) utility of S (respectively, $-L$) has success on it.* c_1 *and* c_2 *represents the costs associated with the effort put by the child in the activity and* 1 *and* 0 *are the payoffs he gets from succeeding or failing.*

3.1 Symmetric and Complete Information

We begin our analysis with the benchmark case in which B knows the state of nature. In this case the announcements made by A would not play any role in the analysis and the behavior of B could be represented by a pair of functions $\alpha, \beta \in [0,1]^{\Theta}$, where $\alpha(\theta_i)$ and $\beta(\theta_i)$ represent the probabilities with which B enters and keeps working on the project in state θ_i.
If $\eta = 0$, it is immediate to check that B would participate and keep working on good projects, while he would not initiate low quality ones and would terminate them whenever started. Assumptions 1 and 2 further imply that this kind of behavior is the one that maximizes the sum of agents' material utility. The same would still be true if we assume that B has reference dependent utility. Although the intuition behind this result is straightforward, the formal proof requires to check the dynamic consistency of different strategies and it is put in the Appendix.

Proposition 1 *If B were to know the state, A's announcement would not play any role and the optimal dynamic consistent strategy for B would be* $(a^*(\theta_L), \beta^*(\theta_L), a^*(\theta_H), \beta^*(\theta_H)) = (0,0,1,1)$.
Proof. *See Appendix 7.1.* ∎

Therefore, if B were to share the same information that A has, his optimal behavior would correspond to the one that maximizes the sum of agents' utilities.

3.2 The Model without Reference-Dependent Utility

In this section, we analyze the model assuming that B does not know the actual state of nature and he does not have reference-dependent utility ($\eta = 0$) either. Although under the assumption of symmetric and complete information, the behavior of B maximizes the sum of material utilities, this is no longer the case if we introduce asymmetric information. The reason for this is a two-sided commitment problem. On the one hand, A is unable to commit herself to tell the truth: if B were to believe her announcement, she would have an incentive to lie in state θ_L claiming that the state is θ_H. On the other hand, B is unable to credibly commit to punish A after a lie: since $c_2 > p_L$, B would rather liquidate the project than keep working on it and any threat of continuing working on the project would not be credible.[28] In the particular case in which B cares about material utility only, this lack of commitment power will make impossible for A to convey any credible information concerning the quality of the project and this will prevent B from initiating the project even if the state is θ_H. The following Proposition formalizes this result:

Proposition 2 *Suppose $\eta = 0$. Then under Assumptions 1 and 2, A's announcement will not affect the participation of B in the project. Furthermore, in the unique equilibrium of the game $\alpha(m) = 0$, $\beta(m, \theta_L) = 0$ and $\beta(m, \theta_L) = 1$ for each message $m \in M$.*

Proof. *Since $p_L < c_2 < 1$, if B were to participate in the project, he would stop working on it if the state were θ_L and continue working if the state were θ_H regardless of A's announcement. Thus, if B were to participate in the project, A would get a payoff $S + G$ in state θ_H and G in state θ_L. Since $G + S > G > 0$, we conclude that the message sent by A cannot affect the probability with which B's participates in the project. Indeed, let $\bar{m}$ be the message associated with the highest probability of participation. If this probability is positive, Assumption 1(iii) implies that B must assign a probability higher than $\frac{1}{2}$ to the state being θ_H. This requires A to send message $\bar{m}$ more often in state θ_H than in state θ_L. By construction, there must exists another message $\underline{m}$ such that the probability B assigns to θ_H upon listening message $\underline{m}$ is lower than $\frac{1}{2}$ (thus message $\underline{m}$ has to be sent in state θ_L more often than in state θ_H). But then Assumption 1(iii) implies that after message $\underline{m}$, B would not participate in the project and, consequently, A would prefer sending message $\bar{m}$ than message $\underline{m}$ in state θ_L. This establishes the required contradiction. The remaining of the proposition follows from Assumptions 1 and 2.* ■

Observe that the lack of commitment power by A follows from the lack of any cost associated with lies. In the next section, we will show that reference-dependent preferences can endogenously introduce this cost by modifying B's attitude toward risk. A different way to attain the same result is to relax the assumption that B can fully observe the state of nature in period 2. We will address this issue in Section 4.1.

3.3 The Model with Reference-Dependent Utility

Now, suppose that B has reference-dependent preferences represented by function $v(. \mid .)$. We will show that under this assumption, a new informative equilibrium arises in which A reveals the truth state of nature. In this fully informative equilibrium, we can assume, without loss of generality, that the set of messages used by A is given by $M = \{\theta_L, \theta_H\}$,[29] and that message

[28] Observe that keep working on the bad project is the only way in which B can punish A for his lie.

[29] Specifically, we can assume that any other message beside θ_L and θ_H will be interpreted by B in the same way of message θ_L.

$m = \theta_i$ is interpreted as "the state of nature is θ_i". Thus information sets in $\mathcal{I}_{\mathcal{M}}$ will be denoted with θ_i, while those in $\mathcal{I}_{\mathcal{M},\Theta}$ will be denoted with (θ_i, θ_j) $(i, j \in \{L, H\})$.
The characterization of the fully informative equilibrium is divided in two steps: we begin assuming that B is certain that A announced the quality of the project sincerely and we derive the optimal dynamic consistent strategy of B under this assumption. Then, given this result, we will characterize the conditions under which A will be willing to reveal the truth. We denote with t^{Tr} the communication strategy followed by A in a fully revealing equilibrium. By definition, $t^{Tr}(\theta_i)[\theta_i] = 1$ and consequently (1) implies that, upon listening to announcement θ_i, B would assign probability 1 to state θ_i. When no confusion arises, we will save on notation omitting the dependency of other elements on t^{Tr}. Observe that in a fully informative equilibrium, B's reference utility will be determined based on $(\alpha(\theta_i), \beta(\theta_i, \theta_i))$ only and will not take into account the behavior that the agent is planning to follow at information set (θ_i, θ_j), $i \neq j$.
The following lemma characterizes the optimal dynamic consistent strategy of agent B under the assumption that A is following strategy t^{Tr}.

Lemma 1 *Let $\eta > 0$. Then under Assumptions 1 and 2, the optimal dynamic consistent strategy at information set $\theta_i \in \mathcal{I}_M$ given t^{Tr} is given by (α^*, β^*), where:*

$$\alpha^*(\theta_i) = \begin{cases} 0 & \text{if } \theta_i = \theta_L \\ 1 & \text{if } \theta_i = \theta_H \end{cases}, \quad \beta^*(\theta_i, \theta_j) = \begin{cases} 0 & \text{if } (\theta_i, \theta_j) = (\theta_L, \theta_L) \\ \beta_{LH}^{Tr} & \text{if } (\theta_i, \theta_j) = (\theta_L, \theta_H) \\ \beta_{HL}^{Tr} & \text{if } (\theta_i, \theta_j) = (\theta_H, \theta_L) \\ 1 & \text{if } (\theta_i, \theta_j) = (\theta_H, \theta_H) \end{cases}$$

and

$$\beta_{HL}^{Tr} = \begin{cases} 1 & \text{if } \lambda > \frac{c_2(1+\eta)}{p_L \eta} - \frac{1}{\eta} \\ x \in [0,1] & \text{if } \lambda = \frac{c_2(1+\eta)}{p_L \eta} - \frac{1}{\eta} \\ 0 & \text{if } \lambda < \frac{c_2(1+\eta)}{p_L \eta} - \frac{1}{\eta} \end{cases}, \quad \beta_{LH}^{Tr} = \begin{cases} 0 & \text{if } \lambda > \frac{1+\eta}{c_2 \eta} - \frac{1}{\eta} \\ x \in [0,1] & \text{if } \lambda = \frac{1+\eta}{c_2 \eta} - \frac{1}{\eta} \\ 1 & \text{if } \lambda < \frac{1+\eta}{c_2 \eta} - \frac{1}{\eta} \end{cases}$$

Proof. *Suppose A announced $m = \theta_H$. Then $\pi(\theta_H) = 1$ and $\beta(\theta_H, \theta_L)$ is irrelevant in determining the reference utility. If B thinks of investing in the project and keeping exerting* *effort were he to find out that the state is indeed* θ_H*, his reference utility would be given by*

$$\tilde{v}_1[x] = \begin{cases} 1 & \text{if } x = 1 \\ \\ 0 & \text{if } x = 0 \end{cases}, \quad \tilde{v}_2[x] = \begin{cases} 1 & \text{if } x = -C \\ \\ 0 & \text{otherwise} \end{cases}$$

With this reference point, this strategy would to a utility of $1 - C$*. Given reference utility 13, this strategy will be dynamically consistent as long as:*[30]

$$1 - C \geq C\eta - \eta\lambda$$

which is always satisfied. Applying a reasoning similar to the one used in Appendix 7.1, we can show that the optimal dynamic consistent strategy at information set θ_H *given* t^{Tr} *is the one mentioned in the statement of the proposition. In particular, given the reference utility induced by* $\alpha^*(\theta_H)$ *and* $\beta^*(\theta_H, \theta_H)$*, the behavior prescribed by the optimal dynamic consistent strategy at information set* (θ_H, θ_L)*, namely* $\beta^*(\theta_H, \theta_L)$*, will be to liquidate the project or keep working on it depending on*

$$-c_1 + c_2\eta - \eta\lambda \lesseqgtr p_L - C - (1 - p_L)\eta\lambda$$

Rearranging terms, we get β_{HL}^{Tr}*.*

Now, suppose that A announced $m = \theta_L$*. Then* $\pi(\theta_L) = 0$ *and* $\beta(\theta_L, \theta_H)$ *is irrelevant in determining the reference utility. Then, it is easy to show that the strategy that prescribes not to participate in the project is dynamic consistent and that the reference utility associated with it is given by a degenerate measure on* 0 *in each dimension. Then, we can follow steps similar to those of Appendix 7.1 to conclude that this is also the reference utility associated with the optimal dynamic consistent strategy. Since the reference utility is given by 12, the optimal* $\beta^*(\theta_L, \theta_L)$ *and* $\beta^*(\theta_L, \theta_H)$ *will be given by:*

$$\beta^*(\theta_L, \theta_L) = \begin{cases} 0 & \text{if } \lambda > \frac{p_L(1+\eta)}{c_2\eta} - \frac{1}{\eta} \\ \\ x \in [0,1] & \text{if } \lambda = \frac{p_L(1+\eta)}{c_2\eta} - \frac{1}{\eta} \\ \\ 1 & \text{if } \lambda < \frac{p_L(1+\eta)}{c_2\eta} - \frac{1}{\eta} \end{cases}$$

and β_{LH}^{Tr} *respectively. The assumption* $p_L < c_2$ *implies* $\frac{p_L(1+\eta)}{c_2\eta} - \frac{1}{\eta} < 1$ *so that the only admissible* $\beta^*(\theta_L, \theta_L)$ *will always be* 0*. This concludes the proof.* ■

[30] Since $-c_1 + c_2\eta - \eta\lambda < C\eta - \eta\lambda$, the most profitable deviation is the one in which B does not join the partnership and the utility associated with this deviation is given by $C\eta - \eta\lambda$.

An immediate corollary of Lemma (1) is that the reference utility associated with the optimal dynamic consistent strategy given t^{Tr} is equal to:

$$\tilde{v}_1[x] = \begin{cases} 1 & \text{if } x = 0 \\ \\ 0 & \text{otherwise} \end{cases}, \quad \tilde{v}_2[x] = \begin{cases} 1 & \text{if } x = 0 \\ \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

at information set $\theta_L \in \mathcal{I}_M$ and to:

$$\tilde{v}_1[x] = \begin{cases} 0 & \text{if } x = 0 \\ \\ 1 & \text{if } x = 1 \end{cases}, \quad \tilde{v}_2[x] = \begin{cases} 1 & \text{if } x = -C \\ \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

at information set $\theta_H \in \mathcal{I}_{\mathcal{M}}$. Intuitively, in a fully informative equilibrium, after listening to announcement θ_H, B assigns probability 1 to the project being of high quality; since in state θ_H, the optimal dynamic consistent strategy is to participate in the project and to keep working on it, the reference utility associated with this announcement will be a degenerate measures on 1 and C in each of the two dimensions of the utility. A similar reasoning justifies the reference utility at information set θ_L. Observe that the optimal dynamic consistent strategy varies with the level of loss aversion. The following Corollary characterizes the optimal behavior if the loss aversion coefficient is high enough.

Corollary 1 *Let $\eta > 0$. Then if Assumption 2 holds, there exists a $\lambda^*(p_L, c_2, \eta)$ such that if $\lambda \geq \lambda^*(p_L, c_2, \eta)$, the optimal dynamic consistent strategy given t^{Tr} prescribes to (i) participate in the project if and only if A announced message $m = \theta_H$, (ii) to keep working on the project regardless of the actual state after message $m = \theta_H$.*[31]

Proof. *The proposition follows immediately from the characterization of the optimal dynamic consistent strategy given in Lemma 1 once we define* $\lambda^*(p_L, c_2, \eta) = \frac{c_2(1+\eta)}{p_L \eta} - \frac{1}{\eta}$. ■

Thus as long as the coefficient of loss aversion is sufficiently high, in the optimal dynamic consistent strategy B will keep working on the project even after finding out that the state is θ_L and that $p_L < c_2$. The explanation for this behavior is as follows. Once B is acquainted with the idea of getting a payoff of 1 (at the cost of incurring an additional cost of c_2), the liquidation of the projects becomes less attractive because it is associated with a relevant psychological loss, while the decision to keep working has the potential to eliminate the

[31] Since node (θ_L, θ_H) and (θ_L, θ_L) never arises on the equilibrium path, we omit specifying the behavior of B, but the full characterization is given in Proposition 1. Observe that

$$\frac{1+\eta}{c_2 \eta} - \frac{1}{\eta} \gtrless \lambda^*(p, c_2, \eta)$$

depeding on whether $p_L \gtrless c_2$.

psychological loss with some positive probability (p_L). To put it differently, if B were to find out that A lied in period 0 (claiming that the project was high quality when it was not), he would face a trade-off between taking the action that maximizes his material utility but is associated with a large psychological loss or forgiving some material utility in the hope of reducing the psychological loss. If the coefficient of loss aversion is sufficiently high, the second option will be more appealing and B will behave as described in Corollary 1.

Observe that the decision of B to keep working on the project is due to a change in his preferences over lotteries determined by the discovery that his future prospects are worse than those promised by A. Indeed, the decision of B on whether to liquidate the project or not can be described as the choice between two lotteries on $\mathbb{R}^2$, where the first dimension represents the outcome-related component and the second one the cost component. *Liquidate* is equivalent to choosing the lottery that delivers outcome $(0, 0)$ for sure,[32] while *Stay* is equivalent to choosing a lottery that delivers outcome $(1, -c_2)$ with probability p_L and payoff $(0, -c_2)$ with probability $(1 - p_L)$. Since $p_L < c_2$, an agent with separable and additive preferences on the two dimensions and no reference dependence would choose the former lottery. However, if the reference utility of the agent is $(1, -c_2)$, negative deviations in the first dimension would be evaluated with weight $\eta\lambda > \eta$, while positive deviation in the second dimension would have weight η. Therefore, as we increase loss aversion, the second lottery becomes more and more attractive and will eventually be preferred to $(0, 0)$. The threshold level for λ at which the change in preference takes place is denoted with $\lambda^* (p_L, c_2, \eta)$. The following Remark summarizes the dependency of this threshold on the other parameters.

Remark 2 $\lambda^* (p_L, c_2, \eta)$ *is increasing in* c_2 *and decreasing in* p_L *and* η.

The previous discussion has shown that, if B believes that A is announcing the quality of the project sincerely and if the degree of loss aversion exceeds a critical threshold, B will react to A's lie by keeping working on the project even after finding out that the project will fail with high probability. The following proposition states that since this type of behavior is sufficiently harmful for A, loss aversion will discipline A and will induce truthtelling.

Proposition 3 *Suppose Assumptions 1 and 2 hold. Then if* $\eta > 0$, *a fully informative equilibrium exists if and only if* $\lambda \geq \lambda^* (p, c_2, \eta)$ *and* $L < -\frac{G+p_L S}{1-p_L}$.

Proof. *Suppose* $\lambda \geq \lambda^* (p, c_2, \eta)$, $L < -\frac{G+p_L S}{1-p_L}$ *and assume that B believes A is following strategy* t^{Tr}. *Consider state* θ_H *first. Then by announcing that the state is* θ_H, *Lemma 1 implies that B would participate in the project and would keep working on it so that the utility of A would be* $G+S$. *On the other hand, if A announces that the state is* θ_L, *Lemma 1 implies that B will not participate in the project and her utility would be* 0. *Since* $G + S > 0$, *A will*

[32] Since at the node where B decides whether to keep working on the project or not c_1 is a sunk cost, it will not play any role in the decision and therefore we omit to include it in the definition of payoffs.

tell the truth in state θ_H. Now consider state θ_L. If A announces the state truthfully, Lemma 1 implies that B will not participate in the project and her utility will be 0*. On the other hand, if A lies and announces that the state is θ_L, by Lemma 1, B will participate in the project and will keep working on it even after finding out that the state is θ_L. In this case A's utility will be given by: $p_L S + (1 - p_L) L + G$. Since $L < -\frac{G+p_L S}{1-p_L}$, the utility from lying is lower than the one from telling the truth. We conclude that in state θ_L, A will announce the type sincerely. Thus a truthful equilibrium exists.*
Suppose that a truthful equilibrium exists. Then the behavior of B is described by Lemma 1. For this to be an equilibrium, in state θ_L, A must send message θ_L instead of message θ_H. By sending message θ_L, he gets utility 0*, while by sending message θ_H he gets utility G if $\lambda < \lambda^* (p, c_2, \eta)$, $\beta(\theta_H, \theta_L)(p_L S + (1 - p_L) L) + G$ if $\lambda = \lambda^* (p, c_2, \eta)$ and $p_L S + (1 - p_L) L + G$ if $\lambda > \lambda^* (p, c_2, \eta)$. Thus, an equilibrium in which A tells the truth can exists only if $\lambda \geq \lambda^* (p, c_2, \eta)$ and $L < -\frac{G+p_L S}{1-p_L}$.* ■

Although the previous analysis has shown the existence of an informative equilibrium in which A reveals his type and B believes her announcement, an uninformative equilibrium also exists. In this equilibrium B ignores A's announcement and consequently A finds optimal sending either message with the same probability regardless of the state confirming B's initial conjecture. To be more precise, let $M = \{\theta_L, \theta_H\}$, but assume that A sends every message with a probability that is independent on the state $t(\theta_i)[\theta_L] = k \in (0, 1)$ $i \in \{L, H\}$. We will denote this strategy with t^{Un}. 1 implies that $\pi\left(m; t^{Un}\right) = \frac{1}{2}$ for every $m \in M$.[33] Furthermore, since $\pi\left(m; t^{Un}\right) = \frac{1}{2}$, whenever $\alpha > 0$, the reference utility of B will be determined by both $\beta(\theta_L)$ and $\beta(\theta_H)$. We begin the analysis of this case by characterizing the optimal dynamic consistent strategy at information set $m \in M$ given t^{Un}.

Proposition 4 *Under Assumptions 1 and 2, if $\eta > 0$, the optimal dynamic consistent strategy at any information set m given t^{Un} is:*

$$(\alpha, \beta(\theta_L), \beta(\theta_H)) = \left(0, 0, \beta_{LH}^{Tr}\right)$$

Furthermore, the reference utility associated with this strategy is a degenerate measure on 0 *for both dimensions.*
Proof. *See Appendix 7.2.* ■

Given the previous proposition, it is immediate to show that A will have no incentive to deviate from strategy t^{Un}.

[33] This follows from Bayes rule under the assumption that $k \in (0, 1]$. If $k = 0$, 1 does not impose any restriction on the updating, but we will still impose $\pi\left(m; t^{Un}\right) = \frac{1}{2}$.

Proposition 5 *If $\eta > 0$ and Assumptions 1 and 2 hold, there exists an uninformative equilibrium in which A follows t^{Un}, B behaves as prescribed by Proposition 4 and beliefs are given by $\pi\left(t^{Un}\right)$.*

Proof. *If A plays an uninformative communication strategy, $\pi(m) = \frac{1}{2}$ for each $m \in M$ and then we have already shown that $\left(0, 0, \beta_{LH}^{Tr}\right)$ is the optimal dynamic consistent strategy for agent B at every m. Furthermore, given that $\pi\left(m, t^{Un}\right) = \frac{1}{2}$ for each m, strategy t^{Un} is trivially optimal for A given $\left(0, 0, \beta_{LH}^{Tr}\right)$.* ■

The previous analysis can be summarized as follows: although in a model in which B is a standard expected utility maximizer ($\eta = 0$) A's is unable to convey any information about the quality of the project, this is no longer true once we allow him to have reference dependent utility ($\eta > 0$). In particular, in this case latter case, a fully informative equilibrium exists. This equilibrium is supported by the threat that B will keep working even on low quality projects. This threat is credible because A's announcement affects not only B's belief concerning the quality of the project, but also his reference utility and, through this second channel, makes B willing to take the risky action in state θ_L. However, as usual in models of communication, this informative equilibrium coexists with another, uninformative equilibrium in which B ignores A's words and A has no incentive to send any information.

4 Extensions and Discussion

In Section 3 we analyze the model under some simplifying assumptions that enable us to focus on the mechanism we were interested in. First of all, we assumed that, upon participating in the project, B learns its quality with certainty and, in this way, we prevented uncertainty and belief distortion to play any role in the second effort decision node. Moreover, Assumption 2 implies that, if B were sure that the project is high quality, his decision to participate would not expose him to any risk or potential loss. In this section we will relax both these assumptions and we will characterize the conditions under which our mechanism still holds. Furthermore, we will also discuss the interaction of reference dependence, loss aversion and communication in more general settings.

4.1 Partial Observability of the State

Until now, we assumed that upon joining the project, agent B learns the true state of the project with certainty. We will now relax this assumption and assume that B finds out the state only with some probability less than 1; note that if B does not find out the true quality of the project, he will base his decision on a belief that could be biased by A's announcement. To capture this idea, we assume that if B enters the project, he learns the true quality of the project only with probability $q \in (0, 1)$; with complementary probability $(1 - q)$, he receives an uninformative signal that does not modify his belief. Formally, consider a set of signals $S = \{\theta_L, 0, \theta_H\}$ and suppose that the conditional probability of receiving signal $x \in S$ if the state is θ_i is given by:

$$\Pr\{s = x \mid \theta_i\} = \begin{cases} q & \text{if } x = \theta_i \\ 1 - q & \text{if } x = 0 \\ 0 & \text{if } x = \theta_j \end{cases}$$

with $i \neq j$. We will keep assuming that Assumptions 1 and 2 hold. The structure of the game is represented in Figure 2.

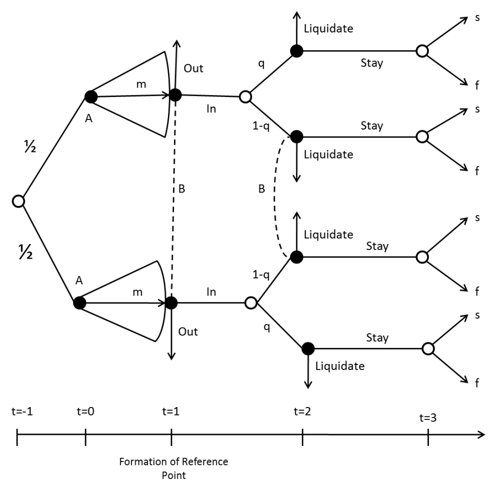


Figure 2: the Model with Partial Observability of the State.

In this framework, the partial observability of the state of nature helps supporting a fully informative equilibrium. The reason is intuitive: suppose that A lied, but B believes she announced the quality of the project truthfully. Since the lie is revealed only with probability q, with complementary probability B will keep working on bad quality projects believing they will succeed with certainty. If q is low enough, A would rather avoid the risk associated with lying regardless of the actual coefficient of loss aversion.

In this framework the behavior of B can be described by a behavioral strategy (α, β_S), where α has the same interpretation as before, and $\beta_S : M \times S \rightarrow [0, 1]$, where $\beta(m, s)$ represents the probability with which B keeps working on the project after that A sent message m, B initiated the project and he received signal s. We will denote with $\pi(m, s; t)$, the probability that B assigns to state θ_H after message m and signal s if he believes that A is following strategy t.

We want to stress that in a fully informative equilibrium, agent B assigns probability 1 to the event that A announced her type truthfully. Thus, upon receiving the uninformative signal concerning the quality of the project, B will keep believing to whatever she announced in period 1. Denoting with t^{Tr} the truthful communication strategy of A, we can write the belief system $\pi\left(t^{Tr}\right)$ as follows:

$$\pi\left(m;t^{Tr}\right)=\begin{cases}1 & m=\theta_H\\ \\ 0 & m=\theta_L\end{cases},\ \pi\left(m,s;t^{Tr}\right)=\begin{cases}1 & \text{if either } s=\theta_H \text{ or } m=\theta_H \text{ and } s=0\\ \\ 0 & \text{if } s=\theta_L\end{cases}$$

The following Proposition characterizes the conditions under which a fully revealing equilibrium exists.

Proposition 6 *Under Assumptions 1 and 2, if agent B finds out the true state only with probability* $q\in(0,1)$, *there exists a truthful equilibrium if either* $\lambda>\lambda^*(p_L,c_2,\eta)$ *or* $L<-\left(\frac{G}{(1-q)(1-p_L)}+\frac{p_LS}{(1-p_L)}\right)$.

Proof. *Suppose that A announced the type truthfully. Then* $\pi(\theta_H)=1$ *and* $\pi(\theta_L)=0$, $\pi(\theta_L,0)=\pi(\theta_H,\theta_L)=0$, $\pi(\theta_H,0)=\pi(\theta_H,\theta_H)=1$. *Consequently we can follow the same reasoning of Proposition 1 to show that the optimal dynamic consistent strategy for agent B at information set* $m=\theta_H$ *is given by* $\alpha(\theta_H)=1$, $\beta(\theta_H,\theta_H)=\beta(\theta_H,0)=1$, $\beta(\theta_H,\theta_L)=\beta^{Tr}_{HL}$, *while the optimal dynamic strategy at information set* $m=\theta_L$ *is given by* $\alpha(\theta_L)=0$, $\beta(\theta_L,\theta_L)=\beta(\theta_L,0)=0$, $\beta(\theta_L,\theta_H)=\beta^{Tr}_{LH}$. *This concludes the characterization of B's optimal dynamic consistent strategy given* t^{Tr}.

Now consider agent A and suppose that the state of nature is θ_H. *By telling the truth she would get a utility of* $G+S$, *while by lying her utility would only be equal to* 0. *Therefore in state* θ_H, *she would always tell the truth. Suppose instead that the state is* θ_L. *By telling the truth, A will get a utility of* 0, *while by lying she would get a utility of*

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

if $\lambda>\lambda^*(p_L,c_2,\eta)$ *and equal to:*

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

if $\lambda<\lambda^*(p_L,c_2,\eta)$.[34] *In the former case, Assumption 2(ii) implies that A would tell the*

[34] The case $\lambda=\lambda^*(p,c_2,\eta)$ is a knife-edge case in which B is free to randomize at information set (θ_H,θ_L). We avoid discussing this case in details since this would not affect the subsequent discussion.

truth. In the latter case, A will tell the truth only if:

$$L < -\left(\frac{p_L S}{(1-p_L)} + \frac{G}{(1-q)(1-p_L)}\right)$$

■

Observe that in the particular case in which $q = 0$, Assumption 2(ii) implies that a fully revealing equilibrium always exist. On the other hand as $q \to 1$, the range of parameters for which the fully revealing equilibrium exists converges to what we would get in the baseline model. For intermediate values of q, a truthful equilibrium exists either if the coefficient of loss aversion is sufficiently high or if the project fails with sufficiently high probability and/or the probability of finding out the true state is high enough.

Although, in both of the contingencies described before the credibility of A comes from the fear that her lie may induce B to keep working on bad quality projects, the mechanism through which this happens is different. If the coefficient of loss aversion is below the threshold level $\lambda^*(p_L, c_2, \eta)$, B would keep working on bad quality projects only if he does not find out the true quality of the project and is under the erroneous assumption that *Stay* is the material utility maximizing strategy. Instead, if $\lambda > \lambda^*(p_L, c_2, \eta)$, B would keep working on the project even if he knows that the true probability of success is p_L. To put it differently, in the first case the behavior of B would be determined by an incorrect belief about the state, while in the latter case, it would be induced by loss aversion and by the desire to reduce the psychological losses associated with giving up.

4.2 Risky High-quality Projects

So far we maintained the assumption that high quality projects succeed with probability 1. Thus, in state θ_H, the project is not only more profitable in expectation, but also riskless: conditional on the state being θ_H, the strategy played by B determines the outcome (and the payoff) without any additional uncertainty. Consequently the characterization of the optimal behavior when the project is high quality does not depend on the coefficient of loss aversion which comes into play only when the project is risky.

Although this assumption is useful to simplify the analysis and to convey the main intuition behind the mechanism that achieves information transmission, the central message of this paper does not depend on it and some interesting insight can be gained by its relaxing it. Thus, we will now allow for some randomness in high-quality projects. There are two ways to introduce this randomness: (i) we can assume that the probability of success in state θ_H is given by $p_H < 1$, or (ii) we can assume that with some exogenous probability B is forced to liquidate the project regardless of its actual quality and to forego the profits associated with it.

We will analyze each case separately and we will show that, under some additional restrictions on parameters that guarantee B's willingness to undertake the project in state θ_H, the main result of this paper is robust. Since, the introduction of randomness in high quality projects, makes the decision of staying out of the project more appealing, it is immediate to see that an uninformative equilibrium in which B ignores A's message and does not enter in the project, still exists regardless of the actual value of η and λ. Furthermore, we can immediately adapt the reasoning developed in Section (3.2) to show that if $\eta = 0$, the only equilibrium is uninformative and involves agent B never participating in the project. Thus, we will focus only on the fully informative equilibrium and we will characterize the conditions under which such an equilibrium exists.

4.2.1 Random Probability of Success

Suppose that the success probability of the project in state θ_i is given by p_i, where $p_i \in (0,1)$ for each $i \in \{L, H\}$ and assume, in line with our interpretation that $p_L < p_H$. In order to analyze the model, we make the following assumptions:

Assumption 3 *(i)* $p_L < c_2 < p_H$*, (ii)* $C < p_H$*, (iii)* $\frac{p_H}{2} < c_1 + \frac{c_2}{2}$

The interpretation of these assumptions is similar to before. In particular, the probability of success is sufficiently high to induce B to start the project and to continue exerting effort on it if the state is θ_H, but sufficiently low to induce him not to participate in the project if he assigns equal probability to both quality of projects.

Suppose $\eta > 0$. Once more, we will characterize the equilibrium in two steps: we start assuming that A follows communication strategy t^{Tr} and we derive the optimal dynamic consistent strategy given t^{Tr}. Then, taking as given the optimal dynamic consistent strategy of B given t^{Tr}, we will show that A has indeed an incentive to announce the state truthfully.

Before beginning our analysis, we define the following functions:

$$\begin{aligned}
\underline{\lambda}\left(p_H, p_L, c_2, \eta\right) &= 1 + \frac{\left(c_2 - p_L\right)\left(1+\eta\right)}{p_H p_L \eta} \\
\bar{\lambda}\left(p_H, c_1, c_2, \eta\right) &= \max\left\{1 + \frac{p_H - C}{p_H\left(1-p_H\right)\eta}, \frac{p_H\left(1+\eta\right)}{C\eta} - \frac{1}{\eta}\right\}
\end{aligned}$$

$\underline{\lambda}\left(p_H, p_L, c_2, \eta\right)$ represents a lower threshold for the degree of loss aversion: if λ exceeded it and if B were under the wrong belief that the project is a high-quality one, he would keep working on projects even if the probability of success is low. On the other hand $\bar{\lambda}\left(p_H, c_1, c_2, \eta\right)$ represents an upper threshold: if λ exceeded it, B's loss aversion would be so high to induce him to avoid projects that can fail with some probability, even though they are high-quality ones.

We are now ready to define the optimal dynamic consistent strategy.

Lemma 2 *Suppose Assumption 3 holds. Then, if $\eta > 0$, the optimal dynamic consistent strategy given t^{Tr} at information set $\theta_H \in \mathcal{I}_M$ is given by:*[35]

$$(\alpha, \beta(\theta_H, \theta_L), \beta(\theta_H, \theta_H)) = \begin{cases} \left(1, \tilde{\beta}_{HL}^{Tr}, 1\right) & \text{if } \lambda \in \left(1, \bar{\lambda}(p_H, c_1, c_2, \eta)\right) \\ \\ \left(0, 0, \tilde{\beta}_{HH}^{Tr}\right) & \text{if } \lambda > \bar{\lambda}(p_H, c_1, c_2, \eta) \end{cases}$$

where

$$\tilde{\beta}_{HL}^{Tr} = \begin{cases} 1 & \text{if } \lambda > \underline{\lambda}(p_H, p_L, c_2, \eta) \\ \\ x \in [0,1] & \text{if } \lambda = \underline{\lambda}(p_H, p_L, c_2, \eta) \\ \\ 0 & \text{if } \lambda < \underline{\lambda}(p_H, p_L, c_2, \eta) \end{cases}, \quad \tilde{\beta}_{HH}^{Tr} = \begin{cases} 0 & \text{if } \lambda > \frac{p_H(1+\eta)}{c_2\eta} - \frac{1}{\eta} \\ \\ x \in [0,1] & \text{if } \lambda = \frac{p_H(1+\eta)}{c_2\eta} - \frac{1}{\eta} \\ \\ 1 & \text{if } \lambda < \frac{p_H(1+\eta)}{c_2\eta} - \frac{1}{\eta} \end{cases}$$

On the other hand, the optimal dynamic consistent strategy at information set $\theta_L \in \mathcal{I}_M$ is given by $\left(0, 0, \tilde{\beta}_{LH}^{Tr}\right)$, where:

$$\tilde{\beta}_{LH}^{Tr} = \begin{cases} 0 & \text{if } \lambda > \frac{p_H(1+\eta)}{c_2\eta} - \frac{1}{\eta} \\ \\ x \in [0,1] & \text{if } \lambda = \frac{p_H(1+\eta)}{c_2\eta} - \frac{1}{\eta} \\ \\ 1 & \text{if } \lambda < \frac{p_H(1+\eta)}{c_2\eta} - \frac{1}{\eta} \end{cases}$$

Proof. *See Appendix 7.3.* ■

An immediate consequence of the previous discussion is that the reference utility at information set θ_H will be

$$\tilde{v}_1[x] = \begin{cases} p_H & \text{if } x = 1 \\ \\ 1 - p_H & \text{if } x = 0 \end{cases}, \quad \tilde{v}_2[x] = \begin{cases} 1 & \text{if } x = -C \\ \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

if $\lambda \in \left(1, \bar{\lambda}(p_H, c_1, c_2, \eta)\right)$ and a degenerate measure on 0 in each dimension if $\lambda > \bar{\lambda}(p_H, c_1, c_2, \eta)$. At information set θ_L, instead, the reference utility of B would be a degenerate measure over 0 in each of the two dimensions.

Thus, whereas initial participation followed by the exertion of effort is always the optimal dynamic consistent strategy when $p_H = 1$, this is no longer true once we introduce some

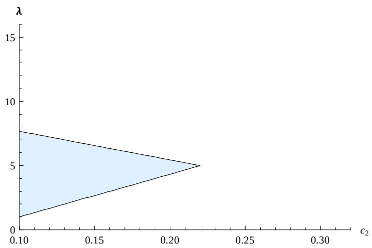
[35] If $\lambda = \bar{\lambda}(p_H, c_1, c_2, \eta)$, any mixture between these two strategies constitute an optimal dynamic consistnet strategy.

randomness in the probability of success. Intuitively, if the high-quality project can fail with some exogenous probability, participation will expose B to some loss; if B is sufficiently loss averse, he will prefer taking the action *Out* and avoid this possibility. Thus, the optimal dynamic consistent reference strategy will entail participation in the project only if the coefficient of loss aversion is not too high. On the other hand, as in Section 3, if the coefficient of loss aversion is high enough, after initial announcement $m = \theta_H$, B would keep working on projects even after finding out that the actual quality of the project is low. Note that the introduction of randomness in the outcome associated with high-quality projects increases the level of loss aversion above which B is willing to incur a material loss in order to decrease the expected psychological loss. The reason behind this result is intuitive: as p_H decreases, the news that the project has high probability of success is associated with a lower probability of getting a outcome-related payoff equal to 1; consequently, the decision of liquidating the project will be less harmful from a psychological point of view and agent B will be less willing to trade off material utility with psychological one.
Therefore, the optimal dynamic consistent strategy is $(1, 1, 1)$ only if the coefficient of loss aversion belongs to an intermediate range of values and in this case, full information transmission would indeed be possible. The next proposition provides conditions under which this range is non-empty and full information transmission is possible.

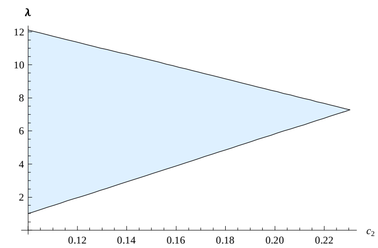
Proposition 7 *Suppose Assumption 3 hold.. Then there exists a threshold level* $\bar{c}_2(p_L, p_H, c_1, \eta) > p_L$ *such that if* $c_2 < \bar{c}_2(p_L, p_H, c_1, \eta)$, $\left[\underline{\lambda}(p_H, p_L, c_2, \eta), \overline{\lambda}(p_H, c_1, c_2, \eta)\right] \neq \emptyset$. *Furthermore a truthful equilibrium in which A's announcements affect B's participation exists if and only if* $\lambda \in \left[\underline{\lambda}(p_H, p_L, c_2, \eta), \overline{\lambda}(p_H, c_1, c_2, \eta)\right]$ *and* $p_L < \frac{-(L+G)}{S-L} < p_H$.
Proof. *See Appendix 7.4.* ■

Thus once we drop the assumption that good quality projects always succeeds, full information transmission is possible only if the cost associated with continued effort is not too high and the coefficient of loss aversion takes intermediate values. In the following remark, we summarize some comparative static results concerning the threshold values of λ and c_2 :

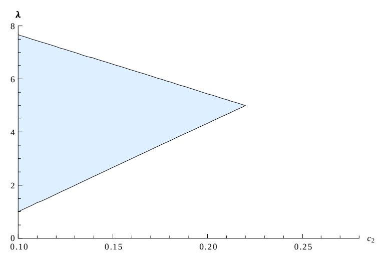
Remark 3 *(i)* $\underline{\lambda}(p_H, p_L, c_2, \eta)$ *is increasing in* c_2 *and decreasing in* η, p_H *and* p_L. *Furthermore* $\underline{\lambda}(1, p_L, c_2, \eta) = \lambda^*(p_L, c_2, \eta) = \frac{c_2(1+\eta)}{p\eta} - \frac{1}{\eta}$.
(ii) $\bar{\lambda}(p_H, c_1, c_2, \eta)$ *is increasing in* p_H, *decreasing in* C *and* η. *Furthermore* $\bar{\lambda}(p_H, c_1, c_2, \eta) \to \infty$ *as* $p_H \to 1$.
(iii) $\bar{c}_2(p_L, p_H, c_1, \eta)$ *is increasing in* p_L. *Furthermore, if* $p_H = 1$, $\bar{c}_2(p_L, p_H, c_1, \eta) = 1 - c_1$ *and Assumption 1 implies* $c_2 < \bar{c}_2(p_L, p_H, c_1, \eta)$.



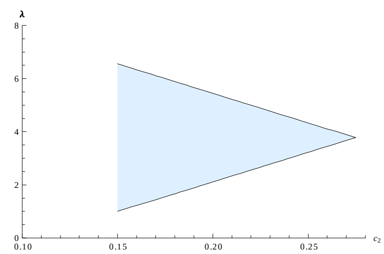
$\eta = 0.5$, $p_H = 0.95$,
$p_L = 0.1$, $c_1 = 0.5$



$\eta = 0.5$, $p_H = 0.9$, $p_L = 0.1$,
$c_1 = 0.5$



$\eta = 0.5$, $p_H = 0.9$, $p_L = 0.1$,
$c_1 = 0.5$



$\eta = 0.5$, $p_H = 0.9$,
$p_L = 0.15$, $c_1 = 0.5$

Figure 3: Set of Parameters for which a Fully Informative Equilibrium Exists.

Figure 3 represents the pairs of loss aversion coefficient λ and cost c_2 for which a fully informative equilibrium exists. The increasing and decreasing function in the figures above represent functions $\underline{\lambda}\left(p_H, p_L, c_2, \eta\right)$ and $\bar{\lambda}\left(p_H, c_1, c_2, \eta\right)$, respectively and the point at which they intersect is the threshold level $\bar{c}_2\left(p_L, p_H, c_1, \eta\right)$ of c_2 below which the range $\left[\underline{\lambda}\left(p_H, p_L, c_2, \eta\right), \bar{\lambda}\left(p_H, c_1, c_2, \eta\right)\right]$ is non empty. From the previous Remark and Figure 3 it is easy to see that the range of admissible values of loss aversion shrinks as c_2 increases and that is empty if $c_2 > \bar{c}_2\left(p_L, p_H, c_1, \eta\right)$.

4.2.2 Random Probability of Liquidation

An alternative way to introduce randomness in high quality projects is to assume that B may experiences some shock that may induce him to liquidate the project even when the probability of succeeding is high. This may happen for different reasons: a sudden need of liquidity that forces liquidation, some personal issue (e.g., health or family problems) may induce a worker to quit his job, the discovery of a different activity may lead a child to stop his current activities.

To be more precise, let $p_H = 1 > p_L$,[36] but assume that with some exogenous probability q independent of the actual quality of the project, a shock may hit B and induce him to liquidate the project regardless of its actual profitability. Suppose that this shock hits B after he initiated the project, but before he has to decide whether to keep working on the project or not. We will further assume that $q < \frac{1}{2}$ and that $1 - q > c_1 + c_2 (1 - q)$. The first assumption capture the idea that the shock is an "exceptional" event, while the second states that the decision of entering in the partnership is the one that maximizes B's material utility if the quality of the project is good. Summing up:

Assumption 4 *(i)* $q < \frac{1}{2}$, *(ii)* $p_L < c_2 < p_H$, *(iii)* $C < 1$, *(iv)* $\frac{(1-q)}{2} < c_1 + \frac{c_2(1-q)}{2}$.

Assumption 5 *(i)* $p_H = 1$.

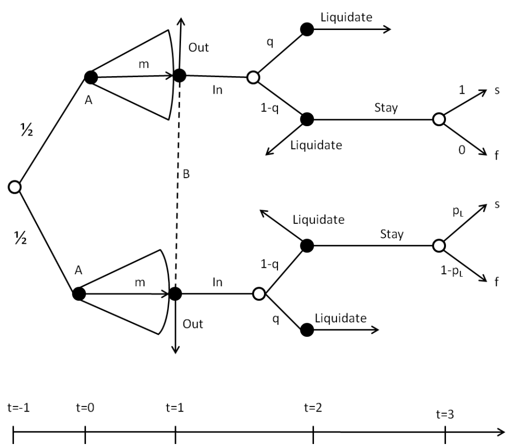


Figure 4: the Model with Exogenous Probability of Liquidation.

Figure 4 represents the structure of the game under these assumptions. As usual, we will define the fully informative strategy of A as t^{Tr} and we will begin our analysis by describing the equilibrium behavior of B under the assumption that A is indeed following communication strategy t^{Tr}. The following proposition provides the full characterization of B's behavior.

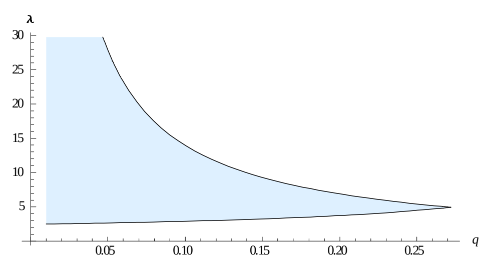
Proposition 8 *Suppose Assumptions 4 and 5 hold. Then we can define a threshold level* $q^* (c_1, c_2, \lambda, \eta)$ *such that if* $q < q^* (c_1, c_2, \lambda, \eta)$, *there exists a non-empty range of loss aversion parameters* $\left[\underline{\underline{\lambda}} (q, c_1, c_2, \eta), \overline{\overline{\lambda}} (q, c_1, c_2, \eta)\right]$ *such that if* $\lambda \in \left[\underline{\underline{\lambda}} (q, c_1, c_2, \eta), \overline{\overline{\lambda}} (q, c_1, c_2, \eta)\right]$,

[36] We assume that the project succeeds with probability 1 in order to focus our attention on the other possible source of randomness, but the two sources of uncertainty could, of course, coexist.

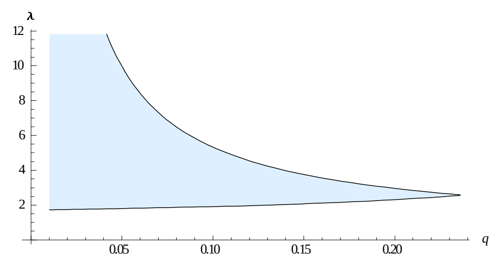
the optimal dynamic consistent strategy given t^{Tr} *is* $(0,0,\beta_{LH})$ *at information set* θ_L *and* $\left(1,\hat{\beta}_{HL}^{Tr},1\right)$ *at information set* θ_H*, where*

$$\hat{\beta}_{LH}^{Tr}=\begin{cases}1 & \text{if } \lambda<\frac{1+\eta}{c_2^L\eta}-\frac{1}{\eta}\\ x\in[0,1] & \text{if } \lambda=\frac{1+\eta}{c_2^L\eta}-\frac{1}{\eta}\\ 0 & \text{if } \lambda>\frac{1+\eta}{c_2^L\eta}-\frac{1}{\eta}\end{cases},\quad \hat{\beta}_{HL}^{Tr}=\begin{cases}1 & \text{if } \lambda>\frac{c_2(1+(1-q)\eta)-p(1+q\eta)}{\eta\left((1-q)p-qc_2^L\right)}\\ x\in[0,1] & \text{if } \lambda=\frac{c_2(1+(1-q)\eta)-p(1+q\eta)}{\eta\left((1-q)p-qc_2^L\right)}\\ 0 & \text{if } \lambda<\frac{c_2(1+(1-q)\eta)-p(1+q\eta)}{\eta\left((1-q)p-qc_2^L\right)}\end{cases}$$

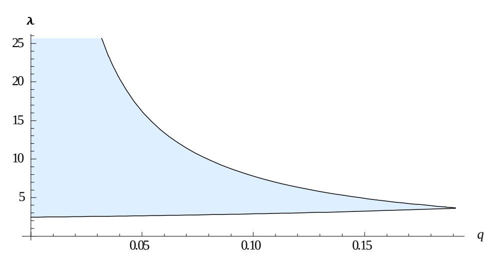
Proof. *See Appendix 7.5.* ■



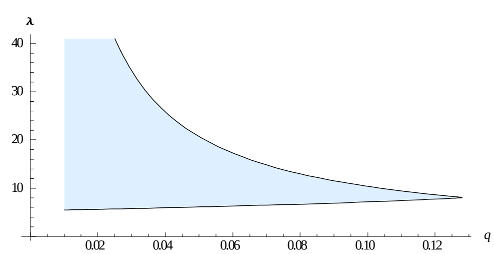
$c_1=0.3$, $c_2=0.2$, $\eta=0.3$, $p_L=0.15$



$c_1=0.3$, $c_2=0.2$, $\eta=0.9$, $p_L=0.15$



$c_1=0.5$, $c_2=0.2$, $\eta=0.3$, $p_L=0.15$



$c_1=0.3$, $c_2=0.3$, $\eta=0.3$, $p_L=0.15$

As in the previous section, loss aversion plays a double role. On the one hand, it discourages B from undertaking in the project because this decision exposes him to some risk. To be more precise, suppose B is certain that the project is high-quality and assume further that he always chooses *Stay* whenever he is not hit by the shock. Then, if we consider the set of lotteries on $\mathbb{R}^2$, action *Out* is equivalent to choosing the lottery that delivers $(0,0)$ for sure, while action *In* is equivalent to choose the lottery that delivers $(0,-c_1)$ with probability q

and $(1, -C)$ with probability $1 - q$. As we increase loss aversion, action In will become less and less appealing since this action may lead to some loss. On the other hand, loss aversion can increase the willingness of B to play $Stay$ when he faces low quality projects, but he was expecting high quality ones. In the following figure, we plot the set of loss aversion coefficients for which the optimal dynamic consistent strategy is $\left(1, \hat{\beta}_{HL}^{Tr}, 1\right)$ as a function of q for different valued of the other parameters.

The next proposition shows that under the conditions mentioned in the previous proposition a fully revealing equilibrium may exist.

Proposition 9 *Suppose Assumptions 4 and 5 hold. Then if* $q < q^* (c_1, c_2, \lambda, \eta)$, *the range* $\left[\underline{\underline{\lambda}} (q, c_1, c_2, \eta), \overline{\overline{\lambda}} (q, c_1, c_2, \eta)\right]$ *is non empty. Furthermore, if* $\lambda \in \left[\underline{\underline{\lambda}} (q, c_1, c_2, \eta), \overline{\overline{\lambda}} (q, c_1, c_2, \eta)\right]$ *and* $L \leq \frac{-G}{(1-q)(1-p_L)} - \frac{p_L S}{(1-p_L)}$, *there exists a fully revealing equilibrium.*

Proof. *The first part of the proposition follows immediately from Proposition 8. Suppose* $\lambda \in \left[\frac{c_2^L - p - qp\eta + (1-q)\eta c_2^L}{\eta\left((1-q)p - qc_2^L\right)}, 1 + \frac{(1-q)\left(1-c_2^L\right) - c_1}{\eta(1-q)q\left(1+c_2^L\right)}\right]$. *We can easily show that A will be sincere when the project is high-quality. Suppose instead that the project is low quality. Then, by announcing the truth A gets 0, while, by lying, she gets* $G + (1-q) p_L S + (1-q)(1-p_L) L$. *Thus A will announce the truth as long as:*

$$L \leq \frac{-G}{(1-q)(1-p_L)} - \frac{p_L S}{(1-p_L)}.$$

■

4.3 Communication and Reference Dependence in Different Settings

So far, we analyzed a model in which the introduction of reference-dependent utility and loss aversion for the Receiver enables credible information transmission when this would not be possible with standard (no reference-dependent) utility. This is possible because Sender's words modify Receiver's reference utility and, through this channel, modify his long run behavior in a way that aligns the interests of the two parties at the initial period.

A natural question is whether the opposite phenomenon can arise, namely whether the introduction of reference dependence and loss aversion can prevent communication instead of facilitating it. This can happen when, without reference dependence, the behavior of the Receiver in the long run would induce the Sender to reveal her information truthfully at the initial period, but the introduction of reference dependence modifies the behavior of the Receiver after a truthful announcement in a way that creates a conflict of interests between the two parties. Consider, for example, a model similar to our, in which there are three types of projects: those that fail for sure as soon as the Receiver participates in them (low quality projects), those which succeed with probability $p < 1$ (medium quality projects) and those which succeed with probability 1 (high quality projects). Assume also that the Sender can distinguish low quality projects from the other two, but cannot further discriminate among projects.[37] Then, with standard utility, if high quality projects are relatively more frequent than medium quality ones, the Receiver will be willing to participate whenever he is able to rule out low quality projects. Therefore, the interests of the two parties are aligned (both want to avoid participation in low quality projects and want to keep working on high quality ones only) and the Sender will reveal her information truthfully. However, for the very same reason we described in Section 2, if the Receiver has reference-dependent utility and is loss averse, the news that the project is not low quality modifies his reference utility in a way that could induce him to keep working on medium quality projects. Since this type of behavior hurts the Sender, she may prefer not to reveal her information and to give up the possibility of inducing participation when she knows that the project is medium or high quality.[38]

Furthermore, in our model we assumed that the Receiver is the only agent with reference dependent attitudes, but the possibility of introducing this type of preferences also for the Sender deserves some comments. To this purpose, it is natural to assume that the reference utility of the Sender is determined at period $t = 0$ and that it is based on her behavior and Receiver's best response. Since the Sender is fully informed about the actual state of nature and moves only once, loss aversion will make her unwilling to take actions that could lead to losses with some positive probability. In our baseline model ($p_H = 1$), these losses arise when the Receiver keeps working on low quality projects that could fail with probability $(1 - p_L)$;[39] therefore, whenever a lie induces the Receiver to keep working on low quality projects, loss aversion will make the Sender even less willing to lie concerning the true quality of the project. If, instead, we allow for randomness in the success probability of high quality projects ($p_H < 1$), inducing participation will be less appealing even in state θ_H, but the analysis we carried out in the previous section will still hold if the loss aversion of the sender is not excessively high.[40]

5 Credibility and Monetary Transfers

Until now, we maintained the assumption that A can announce the quality of the project without having any instrument to make her statement credible. The previous analysis has shown that, in this case, credibility can be established through the interaction of communication,

[37] Thus, if we focus on medium and high quality projects only, we have the same type of interaction as in our baseline model with $p_L = p$, with an *ex-ante* probability of each project possibly different from $\frac{1}{2}$ and with the additional assumption that the Sender is not informed about the true quality of the project.

[38] Details are available upon request.

[39] An alternative channel through which the Sender could experience some loss comes from the decision of the Receiver to randomize at some information set. Since the analysis of this possibility will make the discussion more cumbersome, without affecting the main insight of the present discussion, we will assume that whenever indifferent the Receiver breaks ties by taking a deterministic action.

[40] When loss aversion will exceed this threshold, the Sender will not reveal her information and the Receiver will, consequently, play *Out*.

reference dependence and loss aversion. However, it is not hard to think of situations in which A's announcements are backed by enforceable monetary transfers. Whereas the announcements we considered in previous sections can be interpreted as a cheap and informal way to communicate (e.g., oral communication or nonbinding written statements), these promises represent a more formal type of communication for which A can be held responsible in a court of law. In this section, we will investigate the role played by these monetary transfers and its interaction with loss aversion and reference dependence. This will help us to better understand the two roles that loss aversion plays in our model.

To be more precise, consider the model of Section 4.2.1. To make the comparison with informal communication as direct as possible, we will assume that monetary promises can be enforced at no cost by a third agent (a judge or an independent mediator) and that A does not incur any cost in writing down these enforceable clauses. We start observing that monetary transfers can achieve two different goals. On the one hand, they can enable a Sender with a high quality project to separate himself from one with a low quality one by exploiting differences in the probability with which certain contingencies arise; we will call this the *separation goal.* On the other hand, transfers can decrease the likelihood of incurring a loss by participating in the project and induce B to play In even when, absent any monetary disbursement, the coefficient of loss aversion would prevent him from doing so; we will refer to this objective as to the *participation goal.* To simplify the discussion, we will assume that A is unable to compensate B for the cost associated with effort when the project is bad and, consequently, that the participation goal will be relevant only in state θ_H.

Assumption 6 $G < c_1$.

Enforceable monetary promises will be represented by a function mapping a set of verifiable contingencies into positive real numbers representing the transfers in favor of B; we will denote this mapping with $\kappa : \mathcal{C} \to \mathbb{R}_+$. Thus, the strategy of A will be a function $t : \{\theta_L, \theta_H\} \to \{\theta_L, \theta_H\} \times \mathbb{R}_+^{\mathcal{C}}$ and we will adapt all the definitions of Section 2 in the obvious way.

Now, we can reinterpret Proposition 7 as stating that as long as $c_2 < \bar{c}_2 (p_L, p_H, c_1, \eta)$ and $\lambda \in \left[\underline{\lambda} (p_H, c_1, c_2, \eta), \bar{\lambda} (p_H, c_1, c_2, \eta)\right]$, the separation and participation goal can be attained without using monetary transfers, namely setting $\kappa (.) \equiv 0$. Since informal announcements represent a costless way to induce participation in state θ_H, A will find optimal to use them. In this section, we will analyze what happens when λ does not fall into this interval, but A is allowed to use enforceable monetary transfers. In particular, we will focus on the case in which the set of verifiable contingencies is given by the possible outcome of the project; observe that since $p_H < 1$ these strategies would entail a monetary disbursement with positive probability and, consequently, A would experience a payoff lower than the one achievable with cheap communication.[41]

[41] Given the results in Section 4.2.2, we could include among the set of contingecies the decision of B whether

We begin our analysis with the benchmark case in which the quality of the project is verifiable together with its outcome. In this case, the strategy of A will be a function $t : \{\theta_L, \theta_H\} \to \{\theta_L, \theta_H\} \times \mathbb{R}^4_+$, where $t(\theta_i) = \left(\theta_j, \left(k_s^L, k_f^L\right), \left(k_s^H, k_f^H\right)\right)$ and $\left(k_s^i, k_f^i\right)$ represents the transfers to which A commit herself if the state is θ_i and the project succeeds or fails. It is easy to see that the separation goal can be attained at no cost by playing strategy $\left(\theta_H, (x, x+1), \left(k_s^H, k_f^H\right)\right)$ with $x \geq G$ in state θ_H; the following proposition shows that that the participation goal requires positive transfers when $\lambda > \bar{\lambda}(p_H, c_1, c_2, \eta)$ and characterizes optimal transfers.

Proposition 10 *Suppose that the quality of project and its outcome are verifiable and that Assumption 6 holds. Then, there exists a* $\tilde{\lambda} > 1$ *such that for any* $\lambda \leq \tilde{\lambda}$, *there exists an equilibrium in which: (i) A plays* $\left(\theta_H, (G, 1+G), \left(0, k_f^H(\lambda)\right)\right)$ *if the project is high-quality and* $(\theta_L, (0,0), (0,0))$ *if the project is low-quality; (ii) B assigns probability* 1 *to the project being high quality if A plays* $(\theta_H, (x, 1+x), (y, z))$ *with* $x \geq G$ *and* $x, y \in \mathbb{R}_+$ *and probability* 0 *if she plays something different*[42] *and (iii) the optimal dynamic consistent strategy for B will be* $(1,1,1)$ *at any information set* $(\theta_H, (x, 1+x), (y, z))$ *with* $x \geq G$ *and* $y, z \geq 0$ *and* $(0, 0, \beta((\theta_L, \theta_H), (w, z)))$ *at any other information set* $(., (., .), (w, z))$. *Furthermore:*

$$\beta((\theta_L, \theta_H), (w, z)) = \begin{cases} 1 & \text{if } \lambda < \frac{(p_H(1+w)+(1-p_H)z)(1+\eta)}{C\eta} - \frac{1}{\eta} \\ \\ x \in [0,1] & \text{if } \lambda = \frac{(p_H(1+w)+(1-p_H)z)(1+\eta)}{C\eta} - \frac{1}{\eta} \\ \\ 0 & \text{if } \lambda > \frac{(p_H(1+w)+(1-p_H)z)(1+\eta)}{C\eta} - \frac{1}{\eta} \end{cases}, \tag{15}$$

$k_f^H(\lambda) = 0$ *for any* $\lambda < \bar{\lambda}(p_H, c_1, c_2, \eta)$ *and if* $\lambda \in \left[\bar{\lambda}(p_H, c_1, c_2, \eta), \tilde{\lambda}\right]$

$$k_f^H(\lambda) = \min\left\{\frac{C(1+\lambda\eta)}{(1+\eta)(1-p_H)} - \frac{p_H}{1-p_H}, \frac{(1-p_H)p_H\eta(\lambda-1) + C - p_H}{(1-p_H)(1+p_H\eta(\lambda-1))}\right\}.$$

Proof. *See Appendix 7.6* ■

Proposition 10 states that if the true quality of the project is verifiable, in state θ_H, A can separate himself at no cost by promising to eliminate the loss associated with the project if it were to be a low quality one. However if $\lambda > \bar{\lambda}(p_H, c_1, c_2, \eta)$, A will have to use monetary transfers in order to induce B to participate in the project. The most efficient way to accomplish this goal is to promise a monetary transfer if the project fails ($k_f^H > 0$): indeed, an increase in k_f^H is equivalent to an increase in k_s^H from the material point of view, but

to liquidate the project or not and we would not modify the main insight of this section.

[42] This is not the only class of beliefs that support our equilibrium. Nevertheless this particular beliefs simplify the analysis and do not entail any loss of generality.

has the additional advantage of reducing the psychological loss associated with playing In. Furthermore, as we should expect, the higher the coefficient of loss aversion, the higher this monetary transfer will have to be and the threshold level $\tilde{\lambda}$ represents the point at which inducing participation becomes too costly and A prefers giving up.

Now suppose that the only verifiable events are the outcomes of the project, while the actual qualities are not. In this case the strategy of A can be represented by a triple (m, k_s, k_f) representing the cheap message and the monetary transfers in favor of B if the project succeeds or fails. To simplify the analysis, we impose the following additional assumption:

Assumption 7 *(i)* $c_2 < \bar{c}_2 (p_L, p_H, c_1, \eta)$, *(ii)* $p_L < \frac{1}{2} < p_H$

Assumption 7(i) allows us to focus on the case in which $\underline{\lambda} (p_H, c_1, c_2, \eta) < \bar{\lambda} (p_H, c_1, c_2, \eta)$ and Assumption 7(ii) simplifies the characterization of the optimal monetary transfers when $\lambda < \underline{\lambda} (p_H, c_1, c_2, \eta)$, but the main insight of the following discussion would still hold if it were to be relaxed.

Suppose first that $\lambda > \bar{\lambda} (p_H, c_1, c_2, \eta)$. In this case, since $\lambda \geq \underline{\lambda} (p_H, c_1, c_2, \eta)$, the separation goal can be attained using informal messages in the same way described in Section 4.2.1. Therefore, monetary transfers will be used by A only to induce B to participate in the project; it is not hard to see that this goal can be attained in the same way described in Proposition 10.

Proposition 11 *Suppose that the outcome of the project is verifiable, that Assumption 6 and 7 hold and that* $\lambda > \bar{\lambda} (p_H, c_1, c_2, \eta)$. *Then, there exists a* $\tilde{\lambda} \geq 1$ *such that if* $\lambda \in \left[\bar{\lambda} (p_H, c_1, c_2, \eta), \tilde{\lambda}\right]$ *there exists an equilibrium in which: (i) A plays* $\left(\theta_H, 0, k_f^H (\lambda)\right)$ *if the project is high-quality and* $(\theta_L, 0, 0)$ *if the project is low-quality; (ii) B assigns probability* 1 *to the project being high quality if A played* (θ_H, y, z) *with* $z \geq k_f^H (\lambda)$ *and probability* 0 *if she played something different*[43] *and (iii) the optimal dynamic consistent strategy for B will be* $(1, 1, 1)$ *at any information set* (θ_H, y, z) *with* $z \geq k_f^H (\lambda)$ *and* $(0, 0, \beta ((\theta_L, \theta_H), (w, z)))$ *otherwise.*

Suppose instead that $\lambda < \underline{\lambda} (p_H, c_1, c_2, \eta)$. In this case, if B were sure that the project is high quality, he would always play In. However, given the low value of loss aversion, informal communication is unable to achieve the separation goal. Consequently, a Sender with high quality projects will have to use monetary transfers to separate himself from one with low quality ones.

Proposition 12 *Suppose that the outcome of the project is verifiable, that Assumption 6 and 7 hold and that* $\lambda < \underline{\lambda} (p_H, c_1, c_2, \eta)$. *Then there exists a* $\hat{\lambda} \geq 1$ *such that for any*

[43] Clearly, this is not the only class of beliefs that support our equilibrium. Nevertheless this particular belief simplify the analysis and does not entail any loss of generality.

$\lambda \in \left[\hat{\lambda}, \underline{\lambda}(p_H, c_1, c_2, \eta)\right]$ *there exists an equilibrium in which (i) A follows strategy* $(m, 0, 0)$ *in state* θ_L *and* $\left(m, 0, k_f^{**}(\lambda)\right)$ *in state* θ_H $(m \in \{\theta_L, \theta_H\})$*, (ii) B assigns probability 1 to the project being high quality at any information set* $(m, 0, x)$ *with* $x \geq k_f^{**}(\lambda)$ *and* 0 *otherwise, and (iii) B's optimal dynamic consistent strategy is given by* $(1, 1, 1)$ *at information sets* $(m, 0, x)$ *with* $x \geq k_f^{**}(\lambda)$ *and by* $(0, 0, \beta((\theta_L, \theta_H), (w, z)))$ *at any other information set. Furthermore*

$$k_f^{**}(\lambda) = \frac{(\lambda - 1)\eta p_H p_L - (c_2 - p_L)(1 + \eta)}{p_L(1 + \eta) - (\lambda\eta + 1) + (\lambda - 1)\eta p_H p_L}$$

and $\tilde{\lambda} = 1$*, if* $p_H S + (1 - p_H)\left(L - \frac{c_2 - p_L}{1 - p_L}\right) + G \geq 0$.
Proof. *See Appendix 7.7.* ■

Proposition 12 states that in order to achieve the separation goal, an agent with a high quality project has to induce B to keep exerting effort even after finding out that the project has low probability of success. Once more, the most effective way to attain this goal is to reduce the potential loss associated with the project, namely increasing k_f. In this case, a lower value of λ would imply a low relevance of loss aversion in B's behavior and, consequently, would require higher monetary transfers to induce B to keep working on bad quality projects. Thus, the optimal transfer $k_f^{**}(\lambda)$ will be a decreasing function of λ and for sufficiently low values, A may prefer giving up and avoid inducing participation.

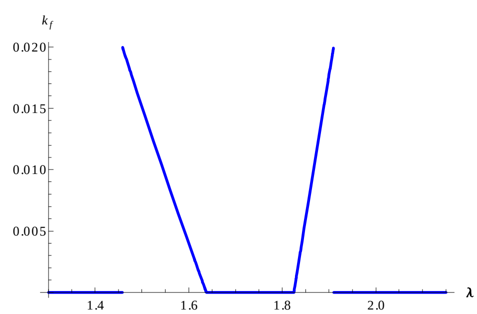


Figure 6: Monetary Transfers as a Function of λ.

An immediate consequence of the previous analysis is that under Assumption 7, there may exist a non-monotonic relationship between the coefficient of loss aversion and the optimal transfers that A has to make in order to induce B to play In in state θ_H: they may be positive for low values of λ, gradually decrease as loss aversion increases, become 0 in the interval $\left[\underline{\lambda}(p_H, c_1, c_2, \eta), \bar{\lambda}(p_H, c_1, c_2, \eta)\right]$ and increase again (and keep doing so) if $\lambda > \bar{\lambda}(p_H, c_1, c_2, \eta)$. Figure 6 represents this pattern for a given set of parameters.

6 Conclusion

In settings where agents have reference-dependent preferences, communication may affect their behavior not only through the change on the probability of states of nature, but also through its effect on agents' prospects and reference points. In the present paper, we exploit this insight to analyze the problem of strategic communication in a model in which an informed agent has an incentive to induce an uninformed party to exert effort.

We show that if the Receiver has standard expected utility, the existence of a conflict of interest between the two parties prevent any credible information transmission. We further show that the introduction of reference dependence and loss aversion may help overcoming this problem as long as the coefficient of loss aversion does not take extreme values: it must be sufficiently low not to prevent the uninformed party from undertaking the project, but also sufficiently high to induce him to modify his behavior in the attempt of avoiding psychological losses. When the loss aversion coefficient is outside this intermediate range, the agent can establish his credibility via monetary transfers which may vary non-monotonically with the degree of loss aversion. In particular, whereas for low degrees of loss aversion transfers are decreasing in loss aversion and are used to separate an agent with high quality projects from one with low quality projects, for high degrees monetary disbursements are increasing and help inducing the uninformed party to undertake risky projects.

The effect that "cheap" announcements can have on agents' expectations and the interaction between these expectations and non-expected utility behaviors is an interesting area for future research. On the one hand, the introduction of multiple senders may shed some light on the circumstances under which informed agents may collude in keeping the receiver ignorant about the true state of nature.[44],[45] Furthermore, the addition of multiple rounds of communication may have nontrivial implications on the sender's behavior: for example, a sender may find himself locked into a chain of lies even though this is suboptimal from an *ex-ante* perspective. Finally, another direction for future research involves the implications of endogenously-determined reference points on contractual design. As shown in Section 5, in a model with loss averse agents, monetary transfers may serve different purposes depending on the actual level of loss aversion; a better analysis of these purposes may help understanding the effect of loss aversion on contractual design.[46]

[44] Intuitively, if the receiver's reference point has been determined under a false announcement, revealing the lie may induce non-optimal behavior thay may hurt the sincere sender.

[45] In Grillo [20] we considered two senders, but we make the extreme assumption that the "types" of the two senders are independent.

[46] A first step in this direction is provided by de Meza and Webb [10].

7 Appendix

7.1 Proof of Proposition 1

Consider state θ_H first. It is immediate to check that strategy $(\alpha(\theta_H), \beta(\theta_H)) = (1,1)$ is dynamic consistent at information set θ_H for any values of parameters and that the total utility associated with this strategy is $1 - C > 0$. Now suppose $\alpha(\theta_H) = 0$; then, even if the strategy were to be dynamic consistent, the total utility of the agent would be 0 and since Assumption 1 implies $1 - C > 0$, we conclude that $\alpha(\theta_H) = 0$ cannot be part of the optimal dynamic consistent strategy in state θ_H. Therefore, in the optimal dynamic consistent strategy $\alpha(\theta_H) > 0$. Suppose $\alpha(\theta_H) > 0$ and $\beta(\theta_H) < 1$. Then the utility associated with this behavior would be

$$-c_1 + \eta c_2 \alpha_H \beta_H - \eta\lambda\alpha_H\beta_H - c_1(1-\alpha)\eta\lambda < 0 < 1 - C$$

and we can conclude that $\alpha(\theta_H) > 0$ and $\beta(\theta_H) < 1$ is not compatible with an optimal dynamic consistent strategy. Finally consider the case $\alpha(\theta_H) \in (0,1)$, $\beta(\theta_H) = 1$. The utility associated with this strategy would be $0 + C\alpha_H\eta - \alpha_H\eta\lambda$ which is lower than $1 - C$. Thus we conclude that the optimal dynamic consistent strategy in state θ_H is given by $(\alpha(\theta_H), \beta(\theta_H)) = (1,1)$.

Now consider state θ_L. It is immediate to see that $(\alpha(\theta_L), \beta(\theta_L)) = (0,0)$ is dynamic consistent at information set θ_L for any value of parameters and that the utility associated with this strategy is 0. Furthermore, strategy $(1, \beta(\theta_L))$ with $\beta(\theta_L) \in [0,1]$ cannot be the optimal dynamic consistent strategy at θ_L. The result is immediate if $\beta(\theta_L) \in \{0,1\}$. If $\beta(\theta_L) \in (0,1)$, the utility associated with this strategy would be:

$$p_L\beta(\theta_L) - c_1 - c_2\beta(\theta_L) - (1-\beta(\theta_L))\beta(\theta_L)c_2\eta(\lambda-1) - p_L\beta(\theta_L)(1-p_L\beta(\theta_L))\eta(\lambda-1)$$

which is negative given Assumption 2. We conclude that the optimal dynamic consistent strategy at information set θ_L, must prescribe $\alpha(\theta_L) < 1$. Suppose that the optimal dynamic strategy at θ_L is $(\alpha(\theta_L), \beta(\theta_L))$ with $\alpha(\theta_L) \in (0,1)$ and $\beta(\theta_L) \in \{0,1\}$. In this case the utility of B would be given by $-p_L - C - (1-\alpha(\theta_L))C\eta\lambda + (1-\alpha(\theta_L))p_L\eta$ if $\beta(\theta_L) = 1$ and by $-c_1 - (1-\alpha(\theta_L))c_1\eta\lambda$ if $\beta(\theta_L) = 0$. Since both these expressions are lower than 0, none of these strategies may be the optimal dynamic consistent one. Finally, suppose that the optimal dynamic consistent strategy at θ_L is $(\alpha(\theta_L), \beta(\theta_L)) \in (0,1) \times (0,1)$. This can be optimal only if all pure strategies give the same utility. Thus we need:

$$\begin{aligned} &0 + \alpha(\theta_L)(1-\beta(\theta_L))c_1\eta + \alpha(\theta_L)\beta(\theta_L)C\eta - \alpha(\theta_L)\beta(\theta_L)p\eta\lambda = \\ &\qquad = -c_1 - \alpha(\theta_L)\beta(\theta_L)p\eta\lambda - (1-\alpha(\theta_L))\eta\lambda c_1 + \alpha(\theta_L)\beta(\theta_L)c_2\eta \end{aligned}$$

and

$$-c_1-\alpha(\theta_L)\beta(\theta_L)p\eta\lambda-(1-\alpha(\theta_L))\eta\lambda c_1+\alpha(\theta_L)\beta(\theta_L)c_2\eta=$$
$$=p-C-\alpha(\theta_L)\beta(\theta_L)p(1-p)\eta\lambda+(1-\alpha(\theta_L)\beta(\theta_L)p)p\eta-(1-\alpha(\theta_L))\eta\lambda C-\alpha_L(1-\beta(\theta_L))c_2\eta\lambda$$

Consider the first equality and observe that it is satisfied as long as:

$$\alpha(\theta_L)(1-\beta(\theta_L))c_1\eta+\alpha(\theta_L)\beta(\theta_L)c_1\eta=-c_1-(1-\alpha(\theta_L))\eta\lambda c_1$$

which cannot be satisfied for any set of parameters. We conclude that the optimal dynamic consistent strategy at θ_L requires $\alpha(\theta_L)=0$. Since in this case, the reference utility is a degenerate measure on 0 for both the monetary and the effort component, one can easily verify that in the optimal dynamic consistent strategy $\beta(\theta_L)=0$.

7.2 Proof of Proposition 4

In an uninformative equilibrium, agent B will not modify his belief concerning the state of nature. Therefore, after announcement $m\in M$, B will assign probability $\frac{1}{2}$ to each of the two possible project's quality. If B thinks of following behavioral strategy $(\alpha,\beta(\theta_L),\beta(\theta_H))$, his reference utility will be given by:

$$\tilde{v}\left(\alpha,\beta;m,t^{Un}\right)=\left(\tilde{v}_1\left(\alpha,\beta;m,t^{Un}\right),\tilde{v}_2\left(\alpha,\beta;m,t^{Un}\right)\right)$$

where:

$$\tilde{v}_1[x]=\begin{cases}\frac{\alpha\beta(\theta_H)}{2}+\frac{\alpha\beta(\theta_L)p_L}{2} & \text{if } x=1\\ \\ 1-\alpha+\frac{\alpha(1-\beta(\theta_H))}{2}+\frac{\alpha(1-\beta(\theta_L)p_L)}{2} & \text{if } x=0\end{cases}\tag{16}$$

in the outcome-related dimension and

$$\tilde{v}_2[x]=\begin{cases}1-\alpha & \text{if } x=0\\ \\ \frac{\alpha}{2}(2-\beta(\theta_H)-\beta(\theta_L)) & \text{if } x=-c_1\\ \\ \frac{\alpha}{2}(\beta(\theta_H)+\beta(\theta_L)) & \text{if } x=-C\end{cases}\tag{17}$$

in the effort dimension.

Assume first that the behavioral strategy followed by B is given by $(0,\beta(\theta_L),\beta(\theta_H))$, $\beta(\theta_L),\beta(\theta_H)\in[0,1]$.[47] In this case the total utility of the agent would be 0. Consider a deviation $(\alpha,\beta(\theta_L),\beta(\theta_H))$

[47] Given that $\alpha=0$, B will never participate to the project and the specification of β_{θ_L} and β_{θ_H} does not affect the formation of the reference utility.

with $\alpha > 0$. In this case given reference utility 16 and 17, the utility of B would be:

$$\alpha\left(\beta\left(\theta_L\right)p_L + \beta\left(\theta_H\right)\right)\left(1+\eta\right) - c_1\alpha\left(1+\eta\lambda\right) - c_2\alpha\left(\beta\left(\theta_L\right)+\beta\left(\theta_H\right)\right)\left(1+\eta\lambda\right)$$

The assumption that $C > \frac{1+p_L}{2}$ implies that this deviation will not be profitable. Furthermore, observe that with reference utility given by 16 and 17, the optimal $\beta\left(\theta_L\right)$ would be 0 and the optimal $\beta\left(\theta_H\right)$ would be given by β_{LH}^{Tr}. Thus $\left(0, 0, \beta_{LH}^{Tr}\right)$ is a dynamic consistent strategy at information set m and delivers a payoff equal to 0.

We will now show that any other dynamic consistent strategy will deliver a lower payoff. Suppose first that B follows a strategy $\left(\alpha, \beta\left(\theta_L\right), \beta\left(\theta_H\right)\right)$ with $\alpha = 1$ is dynamically consistent. Then the reference utility is given by:

$$\tilde{v}_1\left[x\right] = \begin{cases} \frac{\beta(\theta_H)}{2} + \frac{\beta(\theta_L)p_L}{2} & \text{if } x = 1 \\ \\ \frac{1-\beta(\theta_H)}{2} + \frac{1-p_L\beta(\theta_L)}{2} & \text{if } x = 0 \end{cases}, \quad \tilde{v}_2\left[x\right] = \begin{cases} 1 - \frac{\beta(\theta_H)+\beta(\theta_L)}{2} & \text{if } x = -c_1 \\ \\ \frac{\beta(\theta_H)+\beta(\theta_L)}{2} & \text{if } x = -C \end{cases}$$

Suppose $\beta\left(\theta_L\right) > 0$. Then dynamic consistency requires $\beta\left(\theta_H\right) = 1$. To see why note that for $\beta\left(\theta_L\right) > 0$ to be dynamic consistent we need:

$$\begin{aligned} &p_L - C - \left(1 - \frac{\beta\left(\theta_H\right)+\beta\left(\theta_L\right)}{2}\right)c_2\eta\lambda - \left(\frac{\beta\left(\theta_H\right)+\beta\left(\theta_L\right)p_L}{2}\right)\left(1-p_L\right)\eta\lambda + \\ &+\left(1 - \frac{\beta\left(\theta_H\right)+p_L\beta\left(\theta_L\right)}{2}\right)p_L\eta \geq -c_1 + \left(\frac{\beta\left(\theta_H\right)+\beta\left(\theta_L\right)}{2}\right)\eta c_1 - \left(\frac{\beta\left(\theta_H\right)+\beta\left(\theta_L\right)p_L}{2}\right)\eta\lambda \end{aligned}$$

and this inequality implies:

$$\begin{aligned} 1 - C - \left(1 - \frac{\beta\left(\theta_H\right)+\beta\left(\theta_L\right)}{2}\right)c_2\eta\lambda + \left(1 - \frac{\beta\left(\theta_H\right)+p_L\beta\left(\theta_L\right)}{2}\right)\eta \geq \\ \geq -c_1 + \left(\frac{\beta\left(\theta_H\right)+\beta\left(\theta_L\right)}{2}\right)\eta c_1 - \left(\frac{\beta\left(\theta_H\right)+\beta\left(\theta_L\right)p_L}{2}\right)\eta\lambda \end{aligned}$$

Consider strategy $\left(1,1,1\right)$. The utility associated with this strategy is given by:

$$\frac{1+p_L}{2} - C - \left(\frac{1+p_L}{2}\right)\frac{\left(1-p_L\right)}{2}\eta\left(\lambda - 1\right)$$

which is always negative since $C > \frac{1+p_L}{2}$. Similarly, consider strategy $\left(1, \beta\left(\theta_L\right), 1\right)$ and observe that in this case the total utility of B would be:

$$\frac{1+\beta\left(\theta_L\right)p_L}{2} - c_1 - \frac{1+\beta\left(\theta_L\right)}{2}c_2 - \left(\frac{1-p_L^2\beta^2\left(\theta_L\right)}{2}\right)\eta\left(\lambda-1\right) - \frac{1-\beta^2\left(\theta_L\right)}{4}c_2\eta\left(\lambda-1\right)$$

which is once more negative. It is also easy to show that $(1,0,0)$ cannot be the optimal dynamic strategy since the utility associated with this strategy is $-c_1$. The last case to analyze is $(1,0,\beta(\theta_H))$ with $\beta(\theta_H)\in(0,1]$, and the utility associated with this strategy would be:

$$\frac{1}{2}-c_1-\frac{c_2}{2}-\frac{\beta(\theta_H)}{2}\left(1-\frac{\beta(\theta_H)}{2}\right)\eta(\lambda-1)-\left(1-\frac{\beta(\theta_H)}{2}\right)\frac{\beta(\theta_H)}{2}\eta(\lambda-1)$$

which is once again lower than 0. We conclude that the optimal dynamic consistent strategy given t^{Un} cannot entail $\alpha=1$.

Now suppose that B plans to follow strategy $(\alpha,\beta(\theta_L),\beta(\theta_H))$ with $\alpha\in(0,1)$. Reasoning as before, we can show that in a dynamic consistent $\beta(\theta_L)>0$ implies $\beta(\theta_H)=1$. Suppose that $\beta(\theta_L)>0$; then the reference utility is given by:

$$\tilde{v}_1[x]=\begin{cases}\frac{\alpha(1+p_L\beta(\theta_L))}{2} & \text{if } x=1\\ 1-\frac{\alpha(1+p_L\beta(\theta_L))}{2} & \text{if } x=0\end{cases},\quad \tilde{v}_2[x]=\begin{cases}1-\alpha & \text{if } x=0\\ \frac{\alpha}{2}(1-\beta(\theta_L)) & \text{if } x=-c_1\\ \frac{\alpha}{2}(1+\beta(\theta_L)) & \text{if } x=-C\end{cases}$$

The utility B would get from this strategy would be

$$\begin{aligned}&\frac{1+\beta(\theta_L)p_L}{2}-c_1-c_2\left(\frac{1+\beta(\theta_L)}{2}\right)\\ &-\left(\frac{\alpha(1+\beta(\theta_L)p_L)}{2}\right)\left(\frac{1-p_L\beta(\theta_L)}{2}\right)\eta\lambda+\left(1-\frac{\alpha(1+\beta(\theta_L)p_L)}{2}\right)\left(\frac{1+p_L\beta(\theta_L)}{2}\right)\eta\\ &\qquad -(1-\alpha)\left(c_1+\left(\frac{1+\beta(\theta_L)}{2}\right)c_2\right)\eta\lambda-\frac{\alpha\left(1-\beta^2(\theta_L)\right)}{2}\eta(\lambda-1)c_2\end{aligned}$$

which is negative. Thus, no strategy with $\alpha\in(0,1)$ and $\beta(\theta_L)>0$ can be an optimal dynamic consistent strategy. We conclude that if a strategy with $\alpha\in(0,1)$ were to be the optimal dynamic consistent strategy, we would need $\beta(\theta_L)=0$. In this case the reference utility would be:

$$\tilde{v}_1[x]=\begin{cases}\frac{\alpha\beta(\theta_H)}{2} & \text{if } x=1\\ 1-\frac{\alpha\beta(\theta_H)}{2} & \text{if } x=0\end{cases},\quad \tilde{v}_2[x]=\begin{cases}1-\alpha & \text{if } x=0\\ \frac{\alpha(2-\beta(\theta_H))}{2} & \text{if } x=-c_1\\ \frac{\alpha\beta(\theta_H)}{2} & \text{if } x=-C\end{cases}$$

It is immediate to see that a strategy in which $\alpha\in(0,1)$, $\beta(\theta_L)=\beta(\theta_H)=0$ cannot be optimal. Then two cases are possible: $\beta(\theta_H)=1$ and $\beta(\theta_H)\in(0,1)$. Consider the case in which $\alpha \in (0,1), \beta(\theta_L) = 0$ and $\beta(\theta_H) \in (0,1)$. The utility associated with strategy is:

$$\frac{\beta(\theta_H)}{2} - c_1 - \frac{\beta(\theta_H) c_2}{2} - \frac{\alpha\beta(\theta_H)}{2}\left(1 - \frac{\beta(\theta_H)}{2}\right)\eta\lambda + \left(1 - \frac{\alpha\beta(\theta_H)}{2}\right)\frac{\beta(\theta_H)}{2}\eta$$
$$-(1-\alpha)\left(\left(1 - \frac{\beta(\theta_H)}{2}\right)c_1 + \frac{\beta(\theta_H)}{2}C\right)\eta\lambda - \left(1 - \frac{\beta(\theta_H)}{2}\right)\left(\frac{\alpha\beta(\theta_H)}{2}\right)c_2\eta(\lambda - 1)$$

which is negative. Similarly, if $\beta(\theta_H) = 1$, the total utility would be:

$$\frac{1}{2} - c_1 - \frac{c_2}{2} + \left(1 - \frac{\alpha}{2}\right)\frac{1}{2}\eta - \frac{\alpha}{4}\eta\lambda - (1-\alpha)\left(\frac{1}{2}c_1 + \frac{1}{2}C\right)\eta\lambda - \frac{\alpha}{4}c_2\eta(\lambda - 1)$$

and this expression is once again negative.
Thus we conclude that in an uninformative equilibrium the optimal dynamic consistent strategy at information set m is given by $\left(0, 0, \beta_{LH}^{Tr}\right)$ for each $m \in M$ and that the reference utility associated with this strategy is a degenerate measure on 0 for each dimension.

7.3 Proof of Proposition 2

Suppose A announced $m = \theta_H$. Then $\pi\left(\theta_H, t^{Tr}\right) = 1$ and $\beta(\theta_H, \theta_L)$ will be irrelevant in determining the reference utility. If B plans to invest in the project and keep exerting effort were he to find out that the state is indeed θ_H, his reference utility would be given by

$$\tilde{v}_1[x] = \begin{cases} p_H & \text{if } x = 1 \\ \\ 1 - p_H & \text{if } x = 0 \end{cases}, \quad \tilde{v}_2[x] = \begin{cases} 1 & \text{if } x = -C \\ \\ 0 & \text{otherwise} \end{cases}$$

It is immediate to verify that this strategy would lead to a utility of

$$p_H - C - p_H(1 - p_H)\eta(\lambda - 1)$$

This strategy will be dynamically consistent as long as:

$$p_H - C - p_H(1 - p_H)\eta(\lambda - 1) \geq C\eta - p_H\eta\lambda$$

or equivalently

$$\lambda \geq 1 + \frac{(C - p_H)(1 + \eta)}{\eta p_H^2}$$

which is always satisfied since $p_H > C$. Furthermore it is easy to show that, under the reference utility induced by this strategy, the optimal behavior at information set (θ_H, θ_L) is given by:

$$\tilde{\beta}^{Tr}_{HL} = \begin{cases} 1 & \text{if } \lambda > 1 + \frac{(c_2 - p_L)(1+\eta)}{p_H p_L \eta} \\ \\ x \in [0,1] & \text{if } \lambda = 1 + \frac{(c_2 - p_L)(1+\eta)}{p_H p_L \eta} \\ \\ 0 & \text{if } \lambda < 1 + \frac{(c_2 - p_L)(1+\eta)}{p_H p_L \eta} \end{cases}$$

Thus strategy $\left(1, \tilde{\beta}^{Tr}_{HL}, 1\right)$ is dynamic consistent at information set $m = \theta_H$.
Now suppose that some other strategy $(\alpha(m), \beta(m, \theta_L), \beta(m, \theta_H))$ is dynamic consistent at $m = \theta_H$. In what follows, we will denote this strategies omitting their dependency on $m = \theta_H$. In this case the reference utility would be given by:

$$\tilde{v}_1[x] = \begin{cases} \alpha\beta(\theta_H) p_H & \text{if } x = 1 \\ \\ 1 - \alpha\beta(\theta_H) p_H & \text{if } x = 0 \end{cases}, \quad \tilde{v}_2[x] = \begin{cases} 1 - \alpha & \text{if } x = 0 \\ \\ \alpha(1 - \beta(\theta_H)) & \text{if } x = -c_1 \\ \\ \alpha\beta(\theta_H) & \text{if } x = -C \end{cases}$$

Suppose first that $\alpha = 1$ and $\beta(\theta_H) < 1$. For this strategy to be dynamic consistent we need:

$$\beta(\theta_H)\eta c_2 = p_H - c_2 + \beta(\theta_H) p_H^2 \eta\lambda + (1 - \beta(\theta_H) p_H) p_H \eta - (1 - \beta(\theta_H))\eta\lambda c_2$$

Furthermore, we need:

$$\begin{aligned} p_H\beta(\theta_H) - c_1 - c_2\beta(\theta_H) - \beta(\theta_H) p_H (1 - \beta(\theta_H) p_H)\eta(\lambda - 1) - \\ - (1 - \beta(\theta_H))\beta(\theta_H)\eta(\lambda - 1) c_2 \geq c_1\eta + c_2\beta(\theta_H)\eta - \beta(\theta_H) p_H \eta\lambda \end{aligned}$$

or equivalently:

$$\beta(\theta_H)\left(p_H - c_2 + \beta(\theta_H) p_H^2 \eta\lambda + (1 - \beta(\theta_H) p_H) p_H\eta - (1 - \beta(\theta_H))\eta\lambda c_2 - \beta(\theta_H)\eta c_2\right) \geq c_1(1 + \eta)$$

which contradicts the assumption of dynamic consistency.
Now, consider strategy $(0, \beta(\theta_L), \beta(\theta_H))$. Since $\alpha = 0$, $\beta(\theta_L)$ and $\beta(\theta_H)$ do not affect the reference utility of the agent, which is then a degenerate measure on 0 for each of the two

dimensions. This strategy will be dynamic consistent as long as:

$$\lambda > \frac{p_H (1+\eta)}{C\eta} - \frac{1}{\eta}$$

Furthermore, it is immediate to verify that in this case the optimal $\beta(\theta_L)$ is equal to 0 and the optimal $\beta(\theta_H)$ is equal to $\tilde{\beta}^{Tr}_{HH}$. Thus as long as $\lambda > \frac{p_H(1+\eta)}{C\eta} - \frac{1}{\eta}$, $\left(0, 0, \tilde{\beta}^{Tr}_{HH}\right)$ will be a dynamic consistent strategy and will deliver a total utility equal to 0.
Suppose that $\alpha \in (0,1)$ and $\beta_H < 1$. For this to be dynamic consistent, we would need:

$$-c_1 - \alpha\beta_H p_H \eta\lambda - (1-\alpha) c_1 \eta\lambda + \alpha\beta_H c_2 \eta = -\alpha\beta_H p_H \eta\lambda + \alpha (1-\beta_H) c_1 \eta + \alpha\beta_H C\eta$$

which never holds.
Suppose instead that $\alpha \in (0,1)$ and $\beta_H = 1$. In this case the reference utility would be given by:

$$\tilde{v}_1[x] = \begin{cases} \alpha p_H & \text{if } x = 1 \\ 1-\alpha p_H & \text{if } x = 0 \end{cases}, \quad \tilde{v}_2[x] = \begin{cases} 1-\alpha & \text{if } x = 0 \\ \alpha & \text{if } x = -C \end{cases}$$

and for this to be dynamic consistent we would need:

$$0 - \alpha p_H \eta\lambda + \alpha C\eta = p_H - C - \alpha p_H (1-p_H) \eta\lambda + (1-\alpha p_H) p_H \eta - (1-\alpha) C\eta\lambda$$

or equivalently:

$$\lambda = \frac{\left(C - p_H - \eta p_H + \alpha\eta p_H^2 + C\alpha\eta\right)}{\alpha\eta p_H^2 + C\alpha\eta - C\eta}$$

Let $\tilde{\lambda}(p_H, C, \eta, \alpha) = \frac{\left(C - p_H - \eta p_H + \alpha\eta p_H^2 + C\alpha\eta\right)}{\alpha\eta p_H^2 + C\alpha\eta - C\eta}$ and note that $\frac{\partial\tilde{\lambda}(p_H, C, \eta, \alpha)}{\partial\alpha}$ is increasing in α. Since $\tilde{\lambda}(p_H, C, \eta, 1) < 1$, we can conclude that this strategy cannot be dynamic consistent.
The previous analysis can be summarized as follows: if $\lambda < \frac{p_H(1+\eta)}{C\eta} - \frac{1}{\eta}$, the only dynamic consistent strategy at information set $m = \theta_H$ is $\left(1, \tilde{\beta}^{Tr}_{HL}, 1\right)$, while if $\lambda \geq \frac{p_H(1+\eta)}{C\eta} - \frac{1}{\eta}$, there are two dynamic consistent strategies: $\left(1, \tilde{\beta}^{Tr}_{HL}, 1\right)$ and $\left(0, 0, \tilde{\beta}^{Tr}_{HH}\right)$. Therefore, whenever $\lambda \geq \frac{p_H(1+\eta)}{C\eta} - \frac{1}{\eta}$, we will have to deal with the problem of multiple dynamic consistent strategy. One can easily show that the utility associated with continuation strategy $\left(1, \tilde{\beta}^{Tr}_{HL}, 1\right)$ would be higher than the one associated with continuation strategy $\left(0, 0, \tilde{\beta}^{Tr}_{HH}\right)$ as long as:

$$\lambda \leq 1 + \frac{p_H - C}{p_H (1-p_H) \eta}$$

Therefore, the optimal dynamic consistent strategy will be $\left(1, \tilde{\beta}^{Tr}_{HL}, 1\right)$ if

$$\lambda \in \left(1, \max\left\{1 + \frac{p_H - C}{p_H (1 - p_H) \eta}, \frac{p_H (1+\eta)}{C\eta} - \frac{1}{\eta}\right\}\right)$$

and $\left(0, 0, \tilde{\beta}^{Tr}_{HH}\right)$ if $\lambda > \max\left\{1 + \frac{p_H - C}{p_H(1-p_H)\eta}, \frac{p_H(1+\eta)}{C\eta} - \frac{1}{\eta}\right\}$.

Now, suppose that A announced $m = \theta_L$. Then $\pi\left(\theta_L, t^{Tr}\right) = 0$ and $\beta(\theta_L, \theta_H)$ is irrelevant in determining the reference utility. Following the same steps of the case in which $p_H = 1$, we can show that the optimal dynamic consistent strategy at this information set is given by $\left(0, 0, \tilde{\beta}^{Tr}_{LH}\right)$, where:

$$\tilde{\beta}^{Tr}_{LH} = \begin{cases} 0 & \text{if } \lambda > \frac{p_H(1+\eta)}{c_2\eta} - \frac{1}{\eta} \\ x \in [0,1] & \text{if } \lambda = \frac{p_H(1+\eta)}{c_2\eta} - \frac{1}{\eta} \\ 1 & \text{if } \lambda < \frac{p_H(1+\eta)}{c_2\eta} - \frac{1}{\eta} \end{cases}$$

It is also easy to see that, in this case, the reference utility is a degenerate measure on 0 in both dimensions.

7.4 Proof of Proposition 7

Define

$$c_2^*(p_L, p_H, c_1, \eta) = \frac{p_L(1+\eta) - c_1 p_L - \eta p_H p_L}{p_L + (1+\eta)(1-p_H)},$$

$$c_2^{**}(p_L, p_H, c_1, \eta) = \frac{p_L - p_H p_L - c_1 + \sqrt{c_1^2 + p_H^2 p_L^2 + p_L^2 - 2p_H p_L^2 + 2p_L c_1 + 4p_H^2 p_L - 2p_H p_L c_1}}{2}$$

and let

$$\bar{c}_2(p_L, p_H, c_1, \eta) = \max\left\{c_2^*(p_L, p_H, c_1, \eta), c^{**}(p_L, p_H, c_1, \eta)\right\}.$$

Observe that if $c_2 < \bar{c}_2(p_L, p_H, c_1, \eta)$, $\underline{\lambda}(p_H, p_L, c_2, \eta) < \overline{\lambda}(p_H, c_1, c_2, \eta)$. Let $c_2 < \bar{c}_2(p_L, p_H, c_1, \eta)$ and $\lambda \in \left[\underline{\lambda}(p_H, p_L, c_2, \eta), \overline{\lambda}(p_H, c_1, c_2, \eta)\right]$; then the optimal dynamic consistent strategy given t^{Tr} is given by $(1,1,1)$ at information set $m = \theta_H$ and by $\left(0, 0, \tilde{\beta}^{Tr}_{LH}\right)$ at information set $m = \theta_H$.

Let $\lambda \in \left[\underline{\lambda}(p_H, p_L, c_2, \eta), \overline{\lambda}(p_H, c_1, c_2, \eta)\right]$ and $p_L < \frac{-(L+G)}{S-L} < p_H$. Consider the case in which the project is high quality. Then, using Lemma 2, it is immediate to show that by announcing the quality of the project truthfully, A would get a payoff equal to $p_H S + (1-p_H) L + G$, while by announcing that the type is θ_L, his utility would be 0. Thus $p_H > \frac{-(L+G)}{S-L}$ implies that telling the truth would be better than lying. Consider the case in which the project is low quality. Then Lemma 2 implies that by announcing the truth, the utility of A would be 0.

If A were to announce that the state is θ_H instead, his utility would be $p_L S + (1 - p_L) L + G$. $p_L < \frac{-(L+G)}{S-L}$ implies that A will be willing to tell the truth.
Now suppose that a truthful equilibrium in which A's announcement affects B's participation exists. If $\lambda > \overline{\lambda}(p_H, c_1, c_2, \eta)$, Lemma 2 implies that B would not participate in the project regardless of the state and this establishes a contradiction with our assumptions. Thus we need $\lambda \leq \overline{\lambda}(p_H, c_1, c_2, \eta)$. In a truthful equilibrium, after message $m = \theta_H$, B would be certain that the project is high quality and since $\lambda < \overline{\lambda}(p_H, c_1, c_2, \eta)$, Lemma 2 implies he would play strategy $\left(1, \tilde{\beta}^{Tr}_{HL}, 1\right)$. In a truthful equilibrium, if $p_H < \frac{-(L+G)}{S-L}$ and the state were θ_H A would prefer announcing θ_L and prevent participation. Thus for the existence of a truthful equilibrium in which candidates' announcement affect B's participation we need $p_H > \frac{-(L+G)}{S-L}$. Now, we can replicate the same steps in the proof of Proposition 6 to show that $\lambda \geq \underline{\lambda}(p_H, p_L, c_2, \eta)$ and $p_L < \frac{-(L+G)}{S-L}$ are also necessary.

7.5 Proof of Proposition 8

Suppose that B believes that A is following strategy t^{Tr}; then $\pi\left(\theta_L, t^{Tr}\right) = 0$ and $\pi\left(\theta_H, t^{Tr}\right) = 1$. The reference utility of B at information set $m = \theta_i$ would be given by:

$$\tilde{v}_1[x] = \begin{cases} 1 - \alpha(\theta_i)(1-q)\beta(\theta_i, \theta_i)p_i & \text{if } x = 0 \\ \\ \alpha(\theta_i)(1-q)\beta(\theta_i, \theta_i)p_i & \text{if } x = 1 \end{cases},$$

$$\tilde{v}_2[x] = \begin{cases} 1 - \alpha(\theta_i) & \text{if } x = 0 \\ \\ \alpha(\theta_i)(1 - (1-q)\beta(\theta_i, \theta_i)) & \text{if } x = -c_1 \\ \\ \alpha(\theta_i)(1-q)\beta(\theta_i, \theta_i) & \text{if } x = -c_1 - c_2 \end{cases}$$

for $i \in \{L, H\}$.
Reasoning in the usual way, one can show that the optimal dynamic consistent strategy at information set $m = \theta_L$ prescribes $\alpha(\theta_L) = 0$ so that the reference utility at this information set will be a degenerate measure on 0 for each of the two dimensions of the utility function. Given this reference point, it is immediate to verify that $\beta(\theta_L, \theta_L) = 0$ and $\beta(\theta_L, \theta_H) = \hat{\beta}^{Tr}_{LH}$, where

$$\hat{\beta}^{Tr}_{LH} = \begin{cases} 1 & \text{if } \lambda < \frac{1+\eta}{c_2^L \eta} - \frac{1}{\eta} \\ \\ x \in [0,1] & \text{if } \lambda = \frac{1+\eta}{c_2^L \eta} - \frac{1}{\eta} \\ \\ 0 & \text{if } \lambda > \frac{1+\eta}{c_2^L \eta} - \frac{1}{\eta} \end{cases}$$

This conclude the analysis of B at information set $m = \theta_L$.

Now, consider information set $m = \theta_H$. Suppose first that strategy

$$(\alpha(\theta_H), \beta(\theta_H, \theta_L), \beta(\theta_H, \theta_H)) = (1, y, 0, 1, 0)$$

Since $y \in [0, 1]$ does not play any role in determining the reference utility of B, the actual reference utility will be given by:

$$\tilde{v}_1[x] = \begin{cases} q & \text{if } x = 0 \\ & \\ (1-q) & \text{if } x = 1 \end{cases}, \quad \tilde{v}_2[x] = \begin{cases} q & \text{if } x = -c_1 \\ & \\ (1-q) & \text{if } x = -c_1 - c_2 \end{cases}$$

This strategy will be dynamically consistent as long as:

$$1 - c_1 - c_2 + q\eta - qc_2\eta\lambda \geq -c_1 - (1-q)\eta\lambda + (1-q)c_2\eta \tag{18}$$

$$(1-q) - c_1 - (1-q)c_2 - (1-q)q\eta(\lambda - 1)(1 + c_2) \geq -(1-q)\eta\lambda + c_1\eta + (1-q)\eta c_2 \tag{19}$$

and in this case y would be equal to $\tilde{\beta}_{HL}^{Tr}$, where:

$$\tilde{\beta}_{HL}^{Tr} = \begin{cases} 1 & \text{if } \lambda > \frac{c_2(1+(1-q)\eta) - p_L(1+q\eta)}{\eta((1-q)p_L - qc_2)} \\ & \\ x \in [0,1] & \text{if } \lambda = \frac{c_2(1+(1-q)\eta) - p_L(1+q\eta)}{\eta((1-q)p_L - qc_2)} \\ & \\ 0 & \text{if } \lambda < \frac{c_2(1+(1-q)\eta) - p_L(1+q\eta)}{\eta((1-q)p_L - qc_2)} \end{cases}$$

Note that if $q = 0$, inequalities (18)-(19) are satisfied, while if $q = 1$, 19 is not satisfied. Thus there exists a threshold level $\bar{q}\left(c_1, c_2^L, \eta, \lambda\right)$ such that as long as $q < \bar{q}\left(c_1, c_2^L, \eta, \lambda\right)$, $\left(1, \hat{\beta}_{HL}^{Tr}, 1\right)$ is dynamic consistent at information set θ_H given t^{Tr}. Furthermore, inequality 18 is implied by 19 and consequently the strategy will be dynamic consistent as long as inequality 19 is satisfied or equivalently as long as

$$\lambda \geq \frac{(c_1 + c_2(1-q))(1+\eta) - q\eta(1+c_2)(1-q) - (1-q)}{\eta(1-q)(1-q(1+c_2))}$$

Note that the expected utility associated with this strategy is equal to:

$$(1-q) - c_1 - (1-q)c_2 - (1-q)q\eta(\lambda - 1)(1 + c_2)$$

Consider a strategy in which $\alpha(\theta_H) = 1$ and $\beta(\theta_H, \theta_H) < 1$. In this case the reference utility would be given by:

$$\tilde{\upsilon}_1[x] = \begin{cases} 1-\beta(\theta_H,\theta_H)(1-q) & \text{if } x=0 \\ \\ \beta(\theta_H,\theta_H)(1-q) & \text{if } x=1 \end{cases}, \ \tilde{\upsilon}_2[x] = \begin{cases} 1-\beta(\theta_H,\theta_H)(1-q) & \text{if } x=-c_1 \\ \\ \beta(\theta_H,\theta_H)(1-q) & \text{if } x=-c_1-c_2 \end{cases}$$

This strategy is dynamic consistent only if:

$$\begin{aligned} 1-c_1-c_2+(1-\beta(\theta_H,\theta_H)(1-q))\,\eta-(1-\beta(\theta_H,\theta_H)(1-q))\,c_2\eta\lambda = \\ -c_1-\beta(\theta_H,\theta_H)(1-q)\,\eta\lambda+\beta(\theta_H,\theta_H)(1-q)\,c_2\eta \end{aligned}$$

and

$$\begin{aligned} (1-q)\,\beta(\theta_H,\theta_H)-c_1-c_2(1-q)\,\beta(\theta_H,\theta_H)-\beta(\theta_H,\theta_H)(1-q)(1-\beta(\theta_H,\theta_H)(1-q))\,\eta(\lambda-1)- \\ -(1-\beta(\theta_H,\theta_H)(1-q))\,\beta(\theta_H,\theta_H)(1-q)\,\eta(\lambda-1)\,c_2 > \eta\,(c_1+\beta(\theta_H,\theta_H)(1-q)\,c_2-\beta(\theta_H,\theta_H)(1-q)\,\lambda) \end{aligned}$$

and these two conditions are not compatible with each others.
For the very same reason, one can also show that a strategy in which $\alpha(\theta_H)\in(0,1)$ and $\beta(\theta_H,\theta_H)\in(0,1)$ cannot be dynamic consistent: indeed the indifference between *Liquidate* and *Stay* in state θ_L implies that *Out*/leads to a higher utility that *In*.
Now consider the strategy that prescribes $\alpha(\theta_H)=0$. The reference utility associated to this strategy is given by a degenerate measure on 0 for each dimension and it is dynamic consistent as long as:

$$0 \geq (1-q)-c_1-(1-q)\,c_2-c_1\eta\lambda-(1-q)\,\eta\lambda c_2+(1-q)\,\eta$$

$$\Longleftrightarrow$$

$$\lambda \geq \frac{(1-q)(1+\eta)}{(c_1+(1-q)\,c_2)\,\eta}-\frac{1}{\eta}$$

In this case, one can easily show that the dynamic consistent strategy will be given by $\left(0,0,\hat{\beta}^{Tr}_{LH}\right)$. Note that the total utility associated with this strategy will be 0 and observe that the utility associated with strategy $\left(1,\hat{\beta}^{Tr}_{HL},1\right)$ is greater than the one associated with $\left(0,0,\hat{\beta}^{Tr}_{LH}\right)$ as long as:

$$(1-q)-c_1-(1-q)\,c_2-(1-q)\,q\eta(\lambda-1)(1+c_2)\geq 0$$

or equivalently:

$$\lambda \leq 1+\frac{(1-q)(1-c_2)-c_1}{\eta(1-q)\,q\,(1+c_2)}$$

Observe that as long as $q \to 0$, $1 + \frac{(1-q)(1-c_2)-c_1}{\eta(1-q)q(1+c_2)} \to \infty$ and the previous condition is always satisfied.

Consider strategies in which and $\alpha(\theta_H) \in (0,1)$, $\beta(\theta_H, \theta_H) = 1$. Then the reference utility is given by:

$$\tilde{v}_1[x] = \begin{cases} 1 - \alpha(\theta_H)(1-q) & \text{if } x = 0 \\ \\ \alpha(\theta_H)(1-q) & \text{if } x = 1 \end{cases}, \quad \tilde{v}_2[x] = \begin{cases} 1 - \alpha(\theta_H) & \text{if } x = 0 \\ \\ \alpha(\theta_H) q & \text{if } x = -c_1 \\ \\ \alpha(\theta_H)(1-q) & \text{if } x = -c_1 - c_2^L \end{cases}$$

and dynamic consistency would require:[48]

$$(1-q) - c_1 - (1-q) c_2 + (1 - \alpha(1-q))(1-q)\eta - \alpha(1-q) q\eta\lambda - (1-\alpha)(c_1 + (1-q) c_2)\eta\lambda \\ - \alpha q (1-q) c_2 \eta\lambda + \alpha(1-q) q c_2 \eta = -\alpha(1-q)\eta\lambda + \alpha c_1 \eta + \alpha(1-q) c_2 \eta$$

or equivalently:

$$\lambda = \frac{\Big((1+\alpha\eta) c_1 + (1+\alpha\eta) c_2 (1-q) + (1-q)^2 \eta\alpha - q\alpha\eta c_2 (1-q) - (1-q)(1+\eta)\Big)}{(1-q)^2 \alpha\eta - q\alpha\eta c_2 (1-q) - \eta(1-\alpha)(c_1 + c_2(1-q))}$$

Observe that the utility associated with this strategy will always be lower than 0.

The previous analysis implies that as long as $\lambda \leq 1 + \frac{(1-q)(1-c_2)-c_1}{\eta(1-q)q(1+c_2)}$ and

$$\lambda \geq \frac{(c_1 + c_2(1-q))(1+\eta) - (1-q)(1 + q\eta(1+c_2))}{\eta(1-q)(1 - q(1+c_2))}$$

the optimal dynamic consistent strategy is given by $\left(1, \hat{\beta}_{HL}^{Tr}, 1\right)$.

7.6 Proof of Proposition 10

Observe that if A announces $\left(\theta_H, (G, 1+G), \left(k_s^H, k_f^H\right)\right)$ in state θ_H, an agent with low quality projects will (weakly) prefer not to mimic him, because this announcement would lead a payoff equal to 0. Now suppose that B assigns probability 1 to the project being high quality if A offered $((x, 1+x), (y, z))$ with $x \geq G$, $y, z \in \mathbb{R}_+$ and probability 0 if she offered something different. Under these beliefs it is immediate to see that an agent with low quality projects will find weakly optimal to announce $(\theta_L, (0,0), (0,0))$. Thus we conclude that the beliefs described in the proposition are compatible with the strategy we described. Suppose first that $\lambda \leq \bar{\lambda}(p_H, c_1, c_2, \eta)$. Then we can follow the same analysis of Section 4.2.1

[48] In the following expression we omit the dependency of α on θ_H.

to conclude that if B is certain that the project is high quality, he will initiate the project and keep working on it even after finding out that the project is of bad quality; in particular if the project is high quality, he will do that even without monetary transfers and we can conclude that the optimal pair $\left(k_s^H, k_f^H\right)$ for agent A in state θ_H will be $(0,0)$. If instead, B assigns probability 0 the project being high quality, he will not initiate the project ad will play strategy $\left(0,0,\beta_{LH}\left(k_s^H,k_f^H\right)\right)$, where $\beta_{LH}\left(k_s^H,k_f^H\right)$ is given by 15.
Suppose instead that $\lambda > \bar{\lambda}\left(p_H,c_1,c_2,\eta\right)$. The same reasoning as before shows that the separation goal can be accomplished with the same structure of monetary transfers as before. However if $k_s^H = k_f^H = 0$, B will not participate in the project even if he were sure of its high quality. Suppose $\lambda > \bar{\lambda}\left(p_H,c_1,c_2,\eta\right)$; then the following two inequalities are violated:

$$\begin{aligned} 0 &\geq p_H(1+\eta) - \lambda C\eta - C \\ 0 &\geq p_H - C - p_H(1-p_H)\eta(\lambda-1) \end{aligned}$$

and A will look for the cheapest combination of $\left(k_s^H, k_f^H\right)$ that satisfy one of the following two inequalities:

$$\left(p_H\left(1+k_s^H\right)+(1-p_H)k_f^H\right)(1+\eta) - \lambda C\eta - C \geq 0$$
$$p_H\left(1+k_s^H\right)+(1-p_H)k_f^H - C - p_H(1-p_H)\left(1+k_s^H-k_f^H\right)\eta(\lambda-1) \geq 0$$

Without loss of generality, we can assume that the cheapest way to attain this goal is to set $k_s^H = 0$ and $k_f^H > 0$.[49] Therefore, for any $\lambda > \bar{\lambda}\left(p_H,c_1,c_2,\eta\right)$, the optimal monetary transfers is given by $\left(0, k_f^*\right)$, where:

$$k_f^*(\lambda) = \min\left\{\frac{C(1+\lambda\eta)}{(1+\eta)(1-p_H)} - \frac{p_H}{1-p_H}, \frac{(1-p_H)p_H\eta(\lambda-1)+C-p_H}{(1-p_H)(1+p_H\eta(\lambda-1))}\right\}$$

Clearly, this will be feasible as long as the participation constraint of A will not be violated, that is as long as:

$$p_H S + (1-p_H)\left(L - k_f^*(\lambda)\right) + G \geq 0$$

Since k_f^* is increasing in λ, there is a maximal $\hat{\lambda}$ for which $p_H S + (1-p_H)\left(L-k_f^*\right) + G = 0$. The remaining of the proof follows from the same analysis of Section 4.2.1.

[49] In particular, $k_s^H = 0$, $k_f^H > 0$ will be the unique optimal strategy if the second constraint is the one binding, while it will be one of the many optimizers if the first constraint is the binding one.

7.7 Proof of Proposition 12

Observe that in this case an agent with a high quality project will solve the following problem:

$$\max_{k_s,k_f} p_H (S - k_s) + (1 - p_H)(L - k_f) + G$$

subject to:

$$p_H (S - k_s) + (1 - p_H)(L - k_f) + G \geq 0$$

and

$$p_L (1 + k_s) + (1 - p_L) k_f - C - p_H (1 - p_L)(1 + k_s - k_f) \eta\lambda + (1 - p_H) p_L (1 + k_s - k_f) \eta \geq$$
$$\geq -c_1 + c_2\eta - p_H (1 + k_s) \eta\lambda - (1 - p_H) k_f \eta\lambda$$

The first is a participation constraint for agent A. The second constraint guarantees that if B is surprised and faces a low quality project when he was expecting a high quality one, will still be willing to keep working on it. Since this kind of behavior is necessary to prevent an agent with low quality projects from following the same strategy, the second constraint guarantees is the one that guarantees the possibility of achieving the separation goal. Given the assumption that $p_L < \frac{1}{2}$, we can conclude that if we ignore the participation constraint, the previous problem is solved by setting $k_s = 0$ and

$$k_f (\lambda) = \frac{(\lambda - 1) \eta p_H p_L - (c_2 - p_L)(1 + \eta)}{p_L (1 + \eta) - (\lambda\eta + 1) + (\lambda - 1) \eta p_H p_L}$$

which is decreasing in λ (intuitively, the lower λ, the less important the psychological component will be and the more high the material utility will have to be in order to make the "punishment" possible). In particular, if $\lambda = 1$, the optimal k_f would be $\frac{c_2 - p_L}{1 - p_L}$. Thus we conclude that monetary transfers will help inducing participation in good quality projects for any value of $\lambda < \underline{\lambda}(p_H, c_1, c_2, \eta)$ as long as:

$$p_H S + (1 - p_H)\left(L - \frac{c_2 - p_L}{1 - p_L}\right) + G \geq 0$$

If this condition is violated the lowest λ for which we can induce participation will be given by $\tilde{\lambda}$ and will be defined by

$$\frac{(G + p_H S + (1 - p_H) L)}{1 - p_H} = \frac{\left(\tilde{\lambda} - 1\right) \eta p_H p_L - (c_2 - p_L)(1 + \eta)}{p_L (1 + \eta) - \left(\tilde{\lambda}\eta + 1\right) + \left(\tilde{\lambda} - 1\right) \eta p_H p_L}$$

The remaining of the proof is analogous to that of Propositions 2 and 7.

References

[1] Robert J. Aumann and Sergiu Hart, *Long cheap talk*, Econometrica **71** (2002), 1619–1660.

[2] Marco Battaglini, *Multiple referrals and multidimensional cheap talk*, Econometrica **70** (2002), no. 4, 1379–1401.

[3] Pierpaolo Battigalli and Martin Dufwenberg, *Guilt in games*, American Economic Review **97** (2007), no. 2, 170–176.

[4] Pierpaolo Battigalli and Martin Dufwenberg, *Dynamic psychological games*, Journal of Economic Theory **144** (2009), no. 1, 1–35.

[5] Markus K. Brunnermeier and Jonathan A. Parker, *Optimal expectations*, American Economic Review **95** (2005), no. 4, 1092–1118.

[6] Andrew Caplin and John Leahy, *Psychological expected utility theory and anticipatory feelings*, The Quarterly Journal of Economics **116** (2001), no. 1, 55–79.

[7] Gary Charness and Martin Dufwenberg, *Promises and partnership*, Econometrica **74** (2006), no. 6, 1579–1601.

[8] Gary Charness and Martin Dufwenberg, *Participation*, American Economic Review **101** (2011), no. 4, 1211–37.

[9] Vincent P Crawford and Joel Sobel, *Strategic information transmission*, Econometrica **50** (1982), no. 6, 1431–51.

[10] David de Meza and David C. Webb, *Incentive design under loss aversion*, Journal of the European Economic Association **5** (2007), no. 1, 66–92.

[11] Wioletta Dziuda, *Communication with detectable deceit*, Mimeo, January 2011.

[12] Larry G. Epstein, *Living with risk*, Review of Economic Studies **75** (2008), no. 4, 1121–1141.

[13] Joseph Farrell, *Talk is cheap*, American Economic Review **85** (1995), 186–190.

[14] Joseph Farrell and Robert Gibbons, *Cheap talk with two audiences*, American Economic Review **79** (1989), no. 5, 1214–23.

[15] Joseph Farrell and Matthew Rabin, *Cheap talk*, Journal of Economic Perspectives **10** (1996), no. 3, 103–18.

[16] Ernst Fehr, Oliver Hart, and Christian Zehnder, *Contracts as reference points–experimental evidence*, American Economic Review **101** (2011), no. 2, 493–525.

[17] John Geanakoplos, David Pearce, and Ennio Stacchetti, *Psychological games and sequential rationality*, Games and Economic Behavior **1** (1989), no. 1, 60 – 79.

[18] Maria Goltsman, Johannes Hörner, Gregory Pavlov, and Francesco Squintani, *Mediation, arbitration and negotiation*, Journal of Economic Theory **144** (2009), no. 4, 1397–1420.

[19] Jerry R. Green and Nancy L. Stokey, *A two-person game of information transmission*, Journal of Economic Theory **135** (2007), no. 1, 90–104.

[20] Edoardo Grillo, *Electoral speeches and reference dependence: Can politicians' announcement be credible?*, Princeton University, 2011.

[21] Faruk Gul, *A theory of disappointment aversion*, Econometrica **59** (1991), no. 3, pp. 667–686 (English).

[22] Oliver Hart and John Moore, *Contracts as reference points*, The Quarterly Journal of Economics **123** (2008), 1–48.

[23] Maxim Ivanov, *Communication via a strategic mediator*, Journal of Economic Theory **145** (2010), no. 2, 869–884.

[24] Daniel Kahneman, *Reference points, anchors, norms, and mixed feelings*, Organizational Behavior and Human Decision Processes **51** (1992), no. 2, 296–312.

[25] Daniel Kahneman, Jack L Knetsch, and Richard H Thaler, *Experimental tests of the endowment effect and the coase theorem*, Journal of Political Economy **98** (1990), no. 6, 1325–48.

[26] ———, *The endowment effect, loss aversion, and status quo bias: Anomalies*, Journal of Economic Perspectives **5** (1991), no. 1, 193–206.

[27] Daniel Kahneman and Amos Tversky, *Prospect theory: An analysis of decision under risk*, Econometrica **47** (1979), no. 2, 263–91.

[28] Navin Kartik, *Strategic communication with lying costs*, Review of Economic Studies **76** (2009), no. 4, 1359–1395.

[29] Navin Kartik, Marco Ottaviani, and Francesco Squintani, *Credulity, lies, and costly talk*, Journal of Economic Theory **134** (2007), no. 1, 93–116.

[30] Botond Koszegi, *Emotional agency*, The Quarterly Journal of Economics **121** (2006), no. 1, 121–156.

[31] ———, *Utility from anticipation and personal equilibrium*, Economic Theory **44** (2010), no. 3, 415–444.

[32] Botond Koszegi and Matthew Rabin, *A model of reference-dependent preferences*, The Quarterly Journal of Economics **121** (2006), 1133–1165.

[33] ———, *Reference-dependent risk attitudes*, American Economic Review **97** (2007), no. 4, 1047–1073.

[34] ———, *Reference-dependent consumption plans*, American Economic Review **99** (2009), no. 3, 909–36.

[35] Vijay Krishna and John Morgan, *A model of expertise*, The Quarterly Journal of Economics **116** (2001), no. 2, 747–775.

[36] Vijay Krishna and John Morgan, *The art of conversation: eliciting information from experts through multi-stage communication*, Journal of Economic Theory **117** (2004), no. 2, 147–179.

[37] ———, *Cheap talk*, The New Palgrave Dictionary of Economics (Steven N. Durlauf and Lawrence E. Blume, eds.), Palgrave Macmillan, Basingstoke, 2008.

[38] George Loewenstein, *Anticipation and the valuation of delayed consumption*, Economic Journal **97** (1987), no. 387, 666–84.

[39] George Loewenstein and Drazen Prelec, *Anomalies in intertemporal choice: Evidence and an interpretation*, The Quarterly Journal of Economics **107** (1992), no. 2, pp. 573–597 (English).

[40] Marco Ottaviani and Francesco Squintani, *Naive audience and communication bias*, International Journal of Game Theory **35** (2006), no. 1, 129–150.

[41] Matthew Rabin, *Incorporating fairness into game theory and economics*, American Economic Review **83** (1993), no. 5, 1281–1302.

[42] Jonathan Shalev, *Loss aversion equilibrium*, International Journal of Game Theory **29** (2000), no. 2, 269–287.

[43] Robert Sugden, *Reference-dependent subjective expected utility*, Journal of Economic Theory **111** (2003), no. 2, 172–191.

[44] Amos Tversky and Daniel Kahneman, *Loss aversion in riskless choice: A reference-dependent model*, The Quarterly Journal of Economics **106** (1991), no. 4, 1039–61.