From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2012

Essays on Strategic Information Transmission

Grillo, Edoardo

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01xk81jk41f

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

ESSAYS ON STRATEGIC INFORMATION TRANSMISSION

Edoardo Grillo

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF

ECONOMICS

Adviser: Stephen Morris

June 2012

© Copyright by Edoardo Grillo, 2012. All rights reserved.

Abstract

This dissertation analyzes strategic information transmission between informed and uninformed economic agents. Chapter 1 overviews the contents of the dissertation. Chapter 2 and chapter 3 analyze strategic communication under the assumption that the uninformed party has reference-dependent preferences and is loss averse a là Köszegi and Rabin. In particular, chapter 2 studies the link between reference dependence, loss aversion and credible communication in an environment where an uninformed agent (B) has to decide whether to participate or not in a risky project, whose probability of success is known to an informed agent (A) only. We show that reference dependence and loss aversion may give rise to credible information transmission. This happens because inaccurate information has two effects: it leads B to choose the action A prefers in the short run, but it also generates unrealistic expectations that, modifying his reference point, may induce B to take actions in the long run that hurt A. This phenomenon is not possible in a model where B is an expected utility maximizer. In Chapter 3, we use a similar insight to analyze a model of electoral competition in which two parties compete to get the support of a mass of voters. Each party is represented by a politician whose valence is unobservable and can take one of two values: high or low. All voters prefer politicians with high valence, but ideological biases may lead them to vote according to party affiliation. Candidates can make statements concerning their valence; however, if voters are expected utility maximizers, politicians' statements lack any credibility and no information transmission takes place. By introducing reference dependence and loss aversion, information transmission becomes possible. This happens because reference dependence introduces an implicit cost of lying: lies may raise voters' expectations about the candidates' valence and, if detected, may lead them to vote against ideological biases in order to avoid the psychological loss associated with supporting a candidate worse than expected. In this context, we show that the set of parameters under which the fully informative equilibrium exists expands with increases in loss aversion. In chapter 4 we consider a mechanism design problem and we characterize the extent to which an uninformed agent can

use hard evidence to induce informed agents to truthfully reveal their, possibly exclusive, information. More precisely, we study the problem of full implementation in Bayes-Nash equilibrium in environments with incomplete information and hard evidence. We provide a full characterization of the set of implementable social choice functions in economic environments with at least 3 agents, while, in general environments with at least 3 agents, we provide separate necessary and sufficient conditions.

Acknowledgement

I am deeply indebted with my adviser, Stephen Morris, for his encouragement, guidance and patience throughout each step of my doctoral studies at Princeton University; each chapter of this dissertation benefited from countless discussions with him. I am also grateful to Roland Bénabou and Wolfgang Pesendorfer for their comments on various drafts of this project: the final version significantly profited from their advice and insights. I also want to thank all the participants in the Microeconomic Theory Seminar, Microeconomic Theory Student Lunch Seminar and Political Economy Student Workshop at Princeton University and, in particular, Dilip Abreu, Avidit Acharya, Marco Battaglini, Benjamin Brooks, Sylvain Chassang, Faruk Gul, Adam Meirowitz, Juan Ortner, Giridhar Parameswaran, Kristopher Ramsay, Takuo Sugaya and Satoru Takahashi. Finally, I want to thank my wife, Silvia, for her unwavering support during this long journey. Of course, none of the people mentioned above is accountable for the contents of this dissertation, for which I take full responsibility.

Contents

Contents **vi**

1 Introduction **1**

2 Reference Dependence, Risky Projects and Credible Information Transmission **5**

2.1 Communication and Reference Dependence 5

2.1.1 Related Literature 9

2.2 The Model 13

2.2.1 Reference-Dependent Preferences 17

2.3 Reference Dependence and Truth-telling 25

2.3.1 Symmetric and Complete Information 26

2.3.2 The Model without Reference-Dependent Utility 29

2.3.3 The Model with Reference-Dependent Utility 30

2.4 Extensions and Discussion 43

2.4.1 Partial Observability of the State 43

2.4.2 Risky High-quality Projects 47

2.4.3 Communication and Reference Dependence in Other Settings 67

2.5 Credibility and Monetary Transfers 69

2.6 Conclusion 77

3 Reference Dependence and Electoral Speeches **79**

3.1 Electoral Competition and Reference Dependence 79

3.1.1 Related Literature 82

3.2 The Model 85

3.2.1 Reference Dependent Utility 92

3.3 A Simplified Model 101

3.4 Analysis of the General Model 109

3.4.1 Voters with Standard Utility 109

3.4.2 Voters with Reference-Dependent Utility 117

3.5 Conclusion 131

3.6 Proofs 132

3.6.1 Proof of Proposition 19 132

3.6.2 Proof of Proposition 22 144

4 Full Bayesian Implementation with Hard Evidence **149**

4.1 Implementation and Hard Evidence 149

4.1.1 Related Literature 151

4.2 The Model 154

4.3 Necessary Condition 160

4.4 Sufficient Condition 173

4.4.1 Economic Environments 173

4.4.2 General Environments 179

4.5 Conclusion 185

Bibliography **187**

Chapter 1

Introduction

Information is essential for taking sounding decisions and its acquisition plays a key role in many economic and social interactions. Prospective employees try to gather information about the jobs they are offered, while, at the same time, employers try to find out the productivity of their hirings; investors want to find out information concerning the profitability of certain investment opportunities; customers try to learn the quality of certain goods before buying them and so on.

In many environments, information is asymmetrically distributed among agents and those who need it have not direct access to it. For this reason, information transmission has been the focus of a very active research agenda in economics and several models have been proposed to understand when information can be credibly transmitted.[1] In this context, models of cheap talk have received significant attention; communication is cheap if the informed agents can send any message they want without incurring any immediate cost from lying. Besides representing an interesting theoretical benchmark, many real life situations fall (at least to some extent) under these assumptions: for example, in oral communication, the direct cost of lying is often negligible and informed parties are able to misrepresent their private information.

[1] We provide a detailed discussion of this literature in the next chapters.

In this dissertation, we start from the cheap talk benchmark, but we depart from it in two important directions. On the one hand, we allow for the possibility that announcements of the informed agents may modify the expectations of the uninformed ones leading to a change in their preferences; by doing so, we introduce an implicit cost in otherwise cheap announcements: their effect on the reference point of information receivers. On the other hand, we study the boundaries of information acquisition in situations where announcements, although cheap, must be accompanied by some non-counterfeitable evidence that may limit the ability of the informed party to lie. In the remaining of this chapter, we provide a more detailed explanation of the contents of this dissertation.

Chapter 2 considers 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 risky project or not. We assume that agent B evaluates the outcomes he experiences not only in absolute terms, but also with respect to an endogenously determined reference point (reference dependence) and that he suffers from negative deviations from this reference point more than he enjoys equal-size positive ones (loss aversion). In this context, we show that reference dependence and loss aversion modeled a là Kösegi and Rabin may give rise to credible information transmission. This happens because credible announcements, if believed by the uninformed party, play a role in the endogenous determination of the reference point. Thus, inaccurate information, although profitable in the short run, may generate unrealistic expectations that, through the effect of the reference point on B's preferences, may lead B to behave in ways that hurt A in the long run. This phenomenon is not possible in a model where the uninformed agent is an expected utility maximizer. Furthermore, in this chapter, we also investigate the role that reference dependence can play in settings where A can write verifiable contractual clauses involving monetary disbursements to reinforce the credibility of her announcements. We show that these clauses may lead to credible information transmission and that this goal may involve monetary transfers that vary non-monotonically with the degree of loss aversion.

Chapter 3 analyzes the link between reference dependence, loss aversion and cheap communication within a model of electoral competition. In particular, we analyze an environment in which two parties compete to get the support of a mass of voters. Each party is represented by a politician whose valence is unobservable and can take one of two values: high or low. All voters prefer politicians with high valence, but ideological biases may lead them to vote according to party affiliation. Candidates can make statements concerning their valence and they are free to lie at no cost. If voters are expected utility maximizers, politicians' statements lack any credibility and no information transmission takes place; intuitively, since candidates incur no cost from overstating their valence, they will always try to do so destroying their credibility in the electorate. By introducing reference-dependent preferences and loss aversion a là Köszegi and Rabin, we show that full information transmission is attainable. Once more, this happens because candidates' announcements, if credible, modify voters' reference points; because of this change, if voters find out that the candidate of their preferred party pretended to be high valence when he is not, they may decide to vote for the opposing party in order to avoid the psychological loss associated with supporting a candidate worse than expected. We also show that the range of parameters for which the fully informative equilibrium exists enlarges with increases in the degree of loss aversion.

Chapter 4 studies the role played by hard evidence in an information extraction problem. We start assuming that, although agents can send false cheap announcements at no cost, the existence of evidence constraints the information that they can credibly pretend to have. Taking a mechanism design approach, we investigate the circumstances under which an uninformed agent can build incentives schemes (mechanisms) to extract private information that is used to select among a set of possible outcomes (in the mechanism design terminology, to implement a social choice function). To be more precise, we study full implementation of social choice functions in Bayes-Nash equilibrium in environments with non-counterfeitable evidence. In this context, we fully characterize the set of implementable social choice functions in environments where the conflict of interests among agents is sufficiently strong

(economic environments), while, for general environments, we provide a partial characterization by introducing separate necessary and sufficient conditions. In both cases, our proof are constructive: we propose mechanisms that attain full implementation, that is we build mechanisms such that all equilibrium outcomes at each given profile of private information coincide with the one prescribed by the social choice function. These mechanisms work as long as the outcome function that the uninformed agent wants to implement satisfies some requirements relating it with the evidence structure and preferences of agents. This characterization improves on the existing literature, by showing how, in environments with private and possibly exclusive information, the request for evidence provision in support of cheap announcements can prevent opportunistic lies and destroy undesirable collective behaviors.

Chapter 2

Reference Dependence, Risky Projects and Credible Information Transmission

2.1 Communication and Reference Dependence

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 to take a decision that affects the utility of both.[1] Crawford and Sobel (1982) 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

[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.

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 chapter 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

[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.

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 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]

[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 2.4.1.

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; 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 2.4, 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

2.4 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 2.5 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.

2.1.1 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 (1982) 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 of the parties are sufficiently aligned.[7] In the model of this paper,

[5]See, for example, Crawford and Sobel (1982), Farrell and Gibbons (1989), Farrell (1995), Battaglini (2002), Aumann and Hart (2002), Krishna and Morgan (2001),Krishna and Morgan (2004) and the references therein. For surveys on the literature, see Farrell and Rabin (1996) and Krishna and Morgan (2008).

[6]For a different approach see Green and Stokey (2007).

[7]To understand the implications of relaxing the assumption of costless lying, see Kartik (2009) and Kartik, Ottaviani, and Squintani (2007).

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 (2002) 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, Horner, Pavlov, and Squintani (2009) and Ivanov (2010) 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 (2006) 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

[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 (2002).

[10]Goltsman, Horner, Pavlov, and Squintani (2009) deal with an unbiased mediator, while Ivanov (2010) considers the case of a biased one.

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 (2011) 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 (1979). 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, Koszegi and Rabin (2006), Koszegi and Rabin (2007), Koszegi and Rabin (2009) 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 Koszegi and Rabin (2007), Koszegi and Rabin (2009).[15]

[11] See Section 2.4.1 for further details.

[12] See Tversky and Kahneman (1991), Kahneman (1992), Kahneman, Knetsch, and Thaler (1991), Kahneman, Knetsch, and Thaler (1990), Sugden (2003a) and the references therein.

[13] See Shalev (2000) and the references cited below. In a similar way, Gul (1991) provides a theory of disappointment aversion in which a lottery is evaluated based on how negative and positive outcome compare with its endogenously determined certainty equivalent.

[14] See Section 2.2.1 for further details.

[15] See also Kösegi Koszegi (2010). In particular, the "surprise" situations described in Kösegi and Rabin Koszegi and Rabin (2007), Koszegi and Rabin (2009) correspond in our

Koszegi (2006) 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 Grillo (2011a), 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 this 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, Pearce, and Stacchetti (1989) and extended to dynamic settings by Battigalli and Dufwenberg (2009).[17]

Hart and Moore (2008) and Fehr, Hart, and Zehnder (2011) 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 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 (2007) 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 2.5 shows that monetary transfers interact with loss aversion in a nontrivial way and

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 (1987) and Loewenstein and Prelec (1992). For an axiomatic treatment of anticipatory utilities see Caplin and Leahy (2001) and Epstein (2008).

[17] See also Rabin (1993) and Battigalli and Dufwenberg (2007).

this may lead to a non-monotonic relationship between monetary transfers and the coefficient of loss aversion.

Finally, Charness and Dufwenberg (2006) and Charness and Dufwenberg (2011) 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.

2.2 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 2.4.2 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.

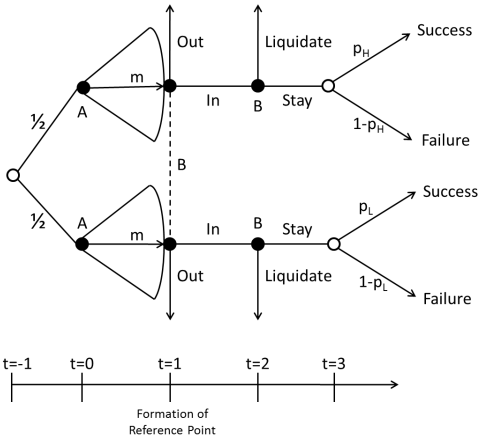


Figure 1: the Game Tree.

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:

[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.

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

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 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

[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.

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 $\alpha : \mathcal{I}_{\mathcal{M}} \to [0, 1]$ and $\alpha(m)$ is the probability with which B chooses In at information set m, while $\beta : \mathcal{I}_{\mathcal{M},\Theta} \to [0, 1]$ 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:

[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.

$$\pi\left(m;\tilde{t}\right)=\frac{\sum\limits_{t:t(\theta_H)=m}\tilde{t}\left[t\right]}{\sum\limits_{\theta\in\{\theta_L,\theta_M\}}\sum\limits_{t:t(\theta)=m}\tilde{t}\left[t\right]} \tag{2.1}$$

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

$$\pi\left(m,\theta_H\right) = 1\ \forall m, \tag{2.2}$$

$$\pi\left(m,\theta_L\right) = 0\ \forall m. \tag{2.3}$$

We will refer to $\left(\pi\left(.;\tilde{t}\right),\pi\left(.,\theta_H\right),\pi\left(.,\theta_H\right)\right)$ 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.

2.2.1 Reference-Dependent Preferences

In this paper we use the utility function introduced by Koszegi and Rabin (2006), Koszegi and Rabin (2007), Koszegi (2010) 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 $u: Z \to \mathbb{R}$. We say that an agent has reference dependent utility if, for any pair of outcomes

[22] Observe that 2.1 and 2.2 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.

$a, r \in Z$, his utility is given by:

$$v\left(a \mid r\right) = u\left(a\right) + \mu\left(u\left(a\right) - u\left(r\right)\right), \tag{2.4}$$

where:

$$\mu\left(x\right) = \eta \cdot \max\left\{0, x\right\} + \eta\lambda \min\left\{0, x\right\} \ \ \forall x \in \mathbb{R} \tag{2.5}$$

with $\eta \in (0, 1)$, and $\lambda > 1$. Thus, the utility of an agent with reference-dependent preferences is represented by a function $v\left(. \mid .\right) : Z \times Z \rightarrow \mathbb{R}$, where the first argument, a, is the actual outcome and the second, r, is the reference outcome. In particular, utility function $v\left(a \mid r\right)$ is the sum of two components: (i) the material or consumption utility represented by utility index $u\left(.\right)$ 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, $u\left(a\right)$, exceeds (respectively, falls short of) the reference utility $u\left(r\right)$, 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, (2.4) and (2.5) 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.

(2.4) and (2.5) can be extended to account for random outcomes $\widetilde{a} \in \Delta\left(Z\right)$ given a fixed reference point $r \in Z$ in the following way:

$$\forall \tilde{a} \in \Delta\left(Z\right), \ v\left(\tilde{a} \mid r\right) = \sum_{a \in Z} v\left(a \mid r\right) \tilde{a}\left(a\right), \tag{2.6}$$

and for random outcomes and reference points as follows:

$$\forall \tilde{a}, \tilde{r} \in \Delta\left(Z\right), \ v\left(\tilde{a} \mid \tilde{r}\right) = \sum_{r \in Z} \sum_{a \in Z} v\left(a \mid r\right) d\tilde{a}\left(a\right) d\tilde{r}\left(r\right), \tag{2.7}$$

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:

$$\forall \tilde{a}, \tilde{r} \in \Delta(Z), \; v(\tilde{a} \mid \tilde{r}) = \sum_{r \in Z} \sum_{a \in Z} \left(\sum_{i=1}^{n} v_i(a_i \mid r_i) \right) d\tilde{a}(a) \, d\tilde{r}(r) \tag{2.8}$$

where for each $i \in \{1, ..., n\}$, $v_i(a_i \mid r_i)$ is defined as in (2.4). Observe that (2.8) 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 Koszegi and Rabin (2006), Koszegi and Rabin (2007), Koszegi and Rabin (2009) 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 $\zeta : D \to Z$ 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, Koszegi and Rabin (2006), Koszegi and Rabin (2007), Koszegi and Rabin (2009) 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 $d' \in D$

$$v(\zeta(d) \mid \zeta(d)) \geq v(\zeta(d') \mid \zeta(d)) \tag{2.9}$$

The second definition of optimality incorporates the effect of deviations on the reference point and selects among all decisions satisfying (2.9) the one that maximizes total utility. Formally, action $d \in D$ is optimal if it satisfies (2.9) and for any other $d' \in D$ satisfying

[23] Conditions that guarantee the existence of these equilibria are provided in Koszegi (2010).

(2.9),

$$v\left(\zeta\left(d\right)\mid\zeta\left(d\right)\right)\geq v\left(\zeta\left(d'\right)\mid\zeta\left(d'\right)\right). \tag{2.10}$$

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 2.1 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:

$$\tilde{u}\left(\alpha,\beta;m,t\right)=\left(\tilde{u}_1\left(\alpha,\beta;m,t\right),\tilde{u}_2\left(\alpha,\beta;m,t\right)\right) \tag{2.11}$$

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:

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

is the random lottery in the cost component.

Similarly the random utility of B when he plays strategy $\left(\alpha,\beta\right)$ at information set $\left(m,\theta_i\right)$ is given by:

$$\tilde{u}\left(\beta;m,\theta_i\right)=\left(\tilde{u}_1\left(\beta;m,\theta_i\right),\tilde{u}_2\left(\beta;m,\theta_i\right)\right),\ i\in\left\{L,H\right\} \tag{2.12}$$

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:

$$\tilde{v}(\alpha, \beta; m, t) = (\tilde{v}_1(\alpha, \beta; m, t), \tilde{v}_2(\alpha, \beta; m, t)), \tag{2.13}$$

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:

$$E\tilde{u} = \sum_{x \in \mathbb{R}} \tilde{u}[x] \cdot x \tag{2.14}$$

and

$$\mu(\tilde{u} - \tilde{v}) = \sum_{x \in \mathbb{R}} \sum_{y \in \mathbb{R}} \mu(\tilde{u}[y] y - \tilde{v}[x] x) \tag{2.15}$$

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:

$$v(\alpha, \beta \mid m, t, \tilde{v}) = E\tilde{u}(\alpha, \beta; m, t) + \mu(\tilde{u}(\alpha, \beta; m, t) - \tilde{v}) \tag{2.16}$$

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

$$v\left(\beta \mid m, \theta_i, \tilde{v}\right) = E\tilde{u}\left(\beta; m, \theta_i\right) + \mu\left(\tilde{u}\left(\beta; m, \theta_i\right) - \tilde{v}\right) \tag{2.17}$$

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 $\tilde{v}\left(\alpha, \beta; m, t\right)$.[24] Formally:

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

$$v\left(\alpha, \beta \mid m, t, \tilde{v}\left(\alpha, \beta; m, t\right)\right) \geq v\left(\hat{\alpha}, \beta \mid m, t, \tilde{v}\left(\alpha, \beta; m, t\right)\right) \tag{2.18}$$

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

$$v\left(\beta \mid m, \theta_i, \tilde{v}\left(\alpha, \beta; m, t\right)\right) \geq v\left(\hat{\beta} \mid m, \theta_i, \tilde{v}\left(\alpha, \beta; m, t\right)\right) \tag{2.19}$$

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

[24] Koszegi (2010) and Koszegi and Rabin (2006), Koszegi and Rabin (2007) and Koszegi and Rabin (2009) 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.

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 (2.9).

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 $\left(\hat{\alpha}, \hat{\beta}\right)$:

$$v\left(\alpha^*, \beta^* \mid m, t, \tilde{v}\left(\alpha^*, \beta^*; m, t\right)\right) \geq v\left(\hat{\alpha}, \hat{\beta} \mid m, t, \tilde{v}\left(\hat{\alpha}, \hat{\beta}; m, t\right)\right) \tag{2.20}$$

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 (2.10).

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

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 2.1, 2.2 and 2.3 given t^*.

[25] In the terminology of Koszegi (2010) and Koszegi and Rabin (2006), Koszegi and Rabin (2007) and Koszegi and Rabin (2009), an optimal dynamic consistent strategy is called *preferred personal equilibrium.*

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

$$\{m : \exists t \text{ such that } t(\theta_H) = m \text{ and } t^*[t] > 0\} \cap$$
$$\cap \{m : \exists t \text{ such that } t(\theta_L) = m \text{ and } t^*[t] > 0\} = \emptyset;$$

- $(t^*, (\alpha^*, \beta^*))$ and $\pi(t^*)$ is uninformative if for every m $\sum_{t:t(\theta_H)=m} t^*[t] = \sum_{t:t(\theta_L)=m} t^*[t]$.

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 (2.18) is equivalent to (2.20). Furthermore (2.18) and (2.19) the definition of equilibrium coincides with that of a Perfect Bayesian Equilibrium.

[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 (2005).

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 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).

2.3 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.

[27] We will relax it in Section 2.4.2.

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.

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.

2.3.1 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: 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:

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

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 $\beta(\theta_L) = 0$. ∎

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.

2.3.2 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 $\bar{m}$

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

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 2.4.1.

2.3.3 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 $m = \theta_i$ 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 (2.1) 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 $(\alpha(\theta_i), \beta(\theta_i, \theta_i))$

[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.

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 Proposition characterizes the optimal dynamic consistent strategy of agent B under the assumption that A is following strategy t^{Tr}.

Proposition 3 Let $\eta > 0$. Then under Assumptions 1 and 2, the optimal dynamic consistent strategy at information set $\theta_i \in I_M$ given t^{Tr} is given by (α^*, β^*), where:

$$\alpha^*(\theta_i) = \begin{cases} 0 & if\ \theta_i = \theta_L \\ \\ 1 & if\ \theta_i = \theta_H \end{cases},$$

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

and

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

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

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

With this reference point, this strategy would to a utility of $1 - C$.

Given reference utility 2.22, 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 the Proof of Proposition 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

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

Rearranging terms, we get β^{Tr}_{HL}.

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 the Proof of Proposition 1 to conclude that this is also the reference utility associated with the optimal dynamic consistent strategy. Since the reference utility is given

[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 $C\eta - \eta\lambda$.

by 2.21, 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. ∎

An immediate corollary of Proposition (3) is that the reference utility associated with the optimal dynamic consistent strategy given t^{Tr} is equal to:

$$\tilde{v}_1[x] = \begin{cases} 1 & \text{if } x = 0 \\ \\ 0 & \text{otherwise} \end{cases}, \ \tilde{v}_2[x] = \begin{cases} 1 & \text{if } x = 0 \\ \\ 0 & \text{otherwise} \end{cases} \tag{2.21}$$

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

$$\tilde{v}_1[x] = \begin{cases} 0 & \text{if } x = 0 \\ \\ 1 & \text{if } x = 1 \end{cases}, \ \tilde{v}_2[x] = \begin{cases} 1 & \text{if } x = -C \\ \\ 0 & \text{otherwise} \end{cases} \tag{2.22}$$

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 3 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 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

[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 3. Observe that

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

depeding on whether $p_L \gtrless c_2$.

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 4 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 3 implies that B would participate in the project and would keep working on it so that the

[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.

utility of A would be $G+S$. On the other hand, if A announces that the state is θ_L, Lemma 3 implies that B will not participate in the project and her utility would be 0. Since $G+S>0$, A will tell the truth in state θ_H. Now consider state θ_L. If A announces the state truthfully, Lemma 3 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 3, 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 3. 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}. 2.1 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}.

[33]This follows from Bayes rule under the assumption that $k\in(0,1]$. If $k=0$, 2.1 does not impose any restriction on the updating, but we will still impose $\pi\left(m;t^{Un}\right)=\frac{1}{2}$.

Proposition 5 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: 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:

$$\tilde{v}_1[x] = \begin{cases} \frac{\alpha\beta(\theta_H)}{2} + \frac{\alpha\beta(\theta_L)p_L}{2} & \text{if } x = 1 \\ 1 - \alpha + \frac{\alpha(1-\beta(\theta_H))}{2} + \frac{\alpha(1-\beta(\theta_L)p_L)}{2} & \text{if } x = 0 \end{cases} \tag{2.23}$$

in the outcome-related dimension and

$$\tilde{v}_2[x] = \begin{cases} 1 - \alpha & \text{if } x = 0 \\ \frac{\alpha}{2}(2 - \beta(\theta_H) - \beta(\theta_L)) & \text{if } x = -c_1 \\ \frac{\alpha}{2}(\beta(\theta_H) + \beta(\theta_L)) & \text{if } x = -C \end{cases} \tag{2.24}$$

in the effort dimension.

Assume first that B's strategy is given by $(0, \beta(\theta_L), \beta(\theta_H))$, $\beta(\theta_L), \beta(\theta_H) \in [0,1]$.[34] In this case the total utility of the agent would be 0. Consider a deviation to strategy $(\alpha, \beta(\theta_L), \beta(\theta_H))$ with $\alpha > 0$. In this case given reference utility 2.23 and 2.24, the utility of B would be:

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

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 2.23 and 2.24, the optimal $\beta(\theta_L)$ would be 0 and the optimal $\beta(\theta_H)$ 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 $(\alpha, \beta(\theta_L), \beta(\theta_H))$ with $\alpha = 1$ is dynamically consistent. Then the reference utility is given by:

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

Suppose $\beta(\theta_L) > 0$. Then dynamic consistency requires $\beta(\theta_H) = 1$. To see why note that

[34] 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.

for $\beta(\theta_L) > 0$ to be dynamic consistent we need:

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

and this inequality implies:

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

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

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

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

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

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:

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

The utility B would get from this strategy would be

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

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:

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

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.

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

Proposition 6 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 5 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.

2.4 Extensions and Discussion

In Section 2.3 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.

2.4.1 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.

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.

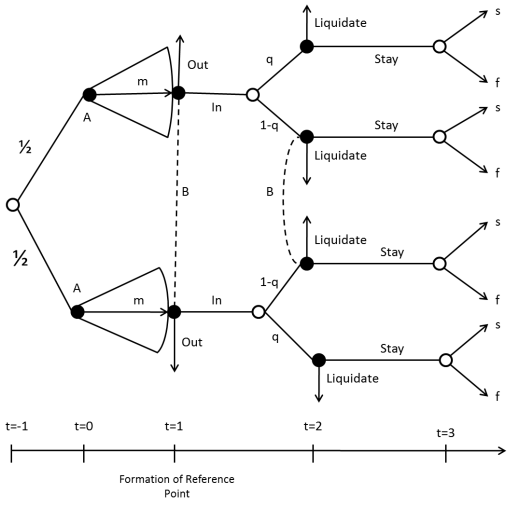


Figure 2: the Model with Partial Observability of the State.

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 $\beta_S : M \times S \to [0,1]$, 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 7 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 $L < -\left(\frac{G}{(1-q)(1-p_L)} + \frac{p_L S}{(1-p_L)}\right)$.

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 3 to show that the optimal dynamic consistent strategy for agent B at information set $m = \theta_H$ is given by $\alpha(\theta_H) = 1$, $\beta(\theta_H, \theta_H) = \beta(\theta_H, 0) = 1$, $\beta(\theta_H, \theta_L) = \beta_{HL}^{Tr}$, while the optimal dynamic strategy at information set $m = \theta_L$ is given by $\alpha(\theta_L) = 0$, $\beta(\theta_L, \theta_L) = \beta(\theta_L, 0) = 0$, $\beta(\theta_L, \theta_H) = \beta_{LH}^{Tr}$. 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

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

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

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

if $\lambda < \lambda^*(p_L, c_2, \eta)$.[35] In the former case, Assumption 2(ii) implies that A would tell the 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.

[35] 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.

2.4.2 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 (2.3.2) 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.

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.

Proposition 8 Suppose Assumption 3 holds. Then, if $\eta > 0$, the optimal dynamic consistent strategy given t^{Tr} at information set $\theta_H \in I_M$ is given by:[36]

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

where

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

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

On the other hand, the optimal dynamic consistent strategy at information set $\theta_L \in I_M$ is

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

given by $\left(0,0,\tilde{\beta}_{LH}^{Tr}\right)$, where:

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

Proof: Suppose A announced $m=\theta_H$. Then $\pi\left(\theta_H,t^{Tr}\right)=1$ and $\beta\left(\theta_H,\theta_L\right)$ 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

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

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

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

This strategy will be dynamically consistent as long as:

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

or equivalently

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

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:

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

Thus strategy $\left(1, \tilde{\beta}_{HL}^{Tr}, 1\right)$ 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:

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

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:

$$p_H\beta(\theta_H) - c_1 - c_2\beta(\theta_H) - \beta(\theta_H) p_H (1 - \beta(\theta_H) p_H)\eta(\lambda - 1) - \\ - (1 - \beta(\theta_H))\beta(\theta_H)\eta(\lambda - 1) c_2 \geq c_1\eta + c_2\beta(\theta_H)\eta - \beta(\theta_H) p_H\eta\lambda$$

or equivalently:

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

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 $\tilde{\beta}_{HH}^{Tr}$. Thus as long as $\lambda > \frac{p_H(1+\eta)}{C\eta} - \frac{1}{\eta}$, $\left(0, 0, \tilde{\beta}_{HH}^{Tr}\right)$ 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:

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

and for this to be dynamic consistent we would need:

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

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 $\tilde{\lambda}\left(p_H, C, \eta, \alpha\right) = \frac{\left(C - p_H - \eta p_H + \alpha\eta p_H^2 + C\alpha\eta\right)}{\alpha\eta p_H^2 + C\alpha\eta - C\eta}$ and note that $\frac{\partial\tilde{\lambda}(p_H, C, \eta, \alpha)}{\partial\alpha}$ is increasing in α. Since $\tilde{\lambda}\left(p_H, C, \eta, 1\right) < 1$, 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 $\left(1, \tilde{\beta}_{HL}^{Tr}, 1\right)$, while if $\lambda \geq \frac{p_H(1+\eta)}{C\eta} - \frac{1}{\eta}$, there are two dynamic consistent strategies: $\left(1, \tilde{\beta}_{HL}^{Tr}, 1\right)$ and $\left(0, 0, \tilde{\beta}_{HH}^{Tr}\right)$. 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 $\left(1, \tilde{\beta}_{HL}^{Tr}, 1\right)$ would be higher than the one associated with continuation strategy $\left(0, 0, \tilde{\beta}_{HH}^{Tr}\right)$ as long as:

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

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

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

and $\left(0, 0, \tilde{\beta}_{HH}^{Tr}\right)$ 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 $\beta\left(\theta_L, \theta_H\right)$ 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 $\left(0, 0, \tilde{\beta}_{LH}^{Tr}\right)$, where:

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

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

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

$$\tilde{v}_1[x] = \begin{cases} p_H & \text{if } x = 1 \\ 1 - p_H & \text{if } x = 0 \end{cases}, \quad \tilde{v}_2[x] = \begin{cases} 1 & \text{if } x = -C \\ 0 & \text{otherwise} \end{cases} \tag{2.25}$$

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 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 2.3, 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 9 Suppose Assumption 3 hold. Then there is a $\bar{c}_2(p_L, p_H, c_1, \eta) > p_L$ 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: Define

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

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

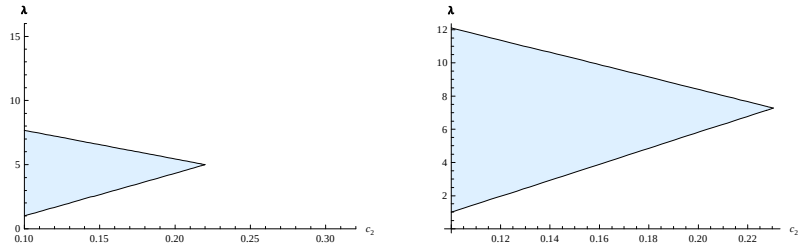
and let

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

Observe that if $c_2 < \bar{c}_2(p_L, p_H, c_1, \eta)$, $\underline{\lambda}(p_H, p_L, c_2, \eta) < \overline{\lambda}(p_H, c_1, c_2, \eta)$.

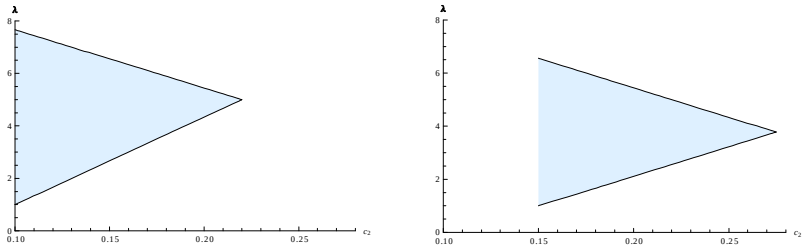
Let $c_2 < \bar{c}_2(p_L, p_H, c_1, \eta)$ and $\lambda \in \left[\underline{\lambda}(p_H, p_L, c_2, \eta), \overline{\lambda}(p_H, c_1, c_2, \eta)\right]$; then the optimal dynamic consistent strategy given t^{Tr} is given by $(1,1,1)$ at information set $m = \theta_H$ and by $\left(0, 0, \tilde{\beta}_{LH}^{Tr}\right)$ at information set $m = \theta_H$. Let $p_L < \frac{-(L+G)}{S-L} < p_H$. Consider the case in which the project is high quality. Then, using Lemma 8, 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 8 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 8 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 8 implies he would play strategy $\left(1, \tilde{\beta}_{HL}^{Tr}, 1\right)$. 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 7 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. ∎



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

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



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

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

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

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 as $p_H \to 1$, $\bar{\lambda}(p_H, c_1, c_2, \eta) \to \infty$.

(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 $c_2 < \bar{c}_2(p_L, p_H, c_1, \eta)$.

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 $\underline{\lambda}(p_H, p_L, c_2, \eta)$ and $\bar{\lambda}(p_H, c_1, c_2, \eta)$, respectively and the point at which they intersect is the threshold level $\bar{c}_2(p_L, p_H, c_1, \eta)$ of c_2 below which the range $\left[\underline{\lambda}(p_H, p_L, c_2, \eta), \bar{\lambda}(p_H, c_1, c_2, \eta)\right]$ 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 $c_2 > \bar{c}_2(p_L, p_H, c_1, \eta)$.

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 $p_H = 1 > p_L$,[37] 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:

[37] 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.

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 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.

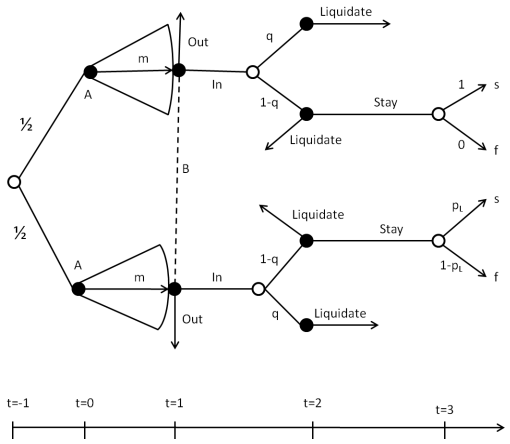


Figure 4: the Model with Exogenous Probability of Liquidation.

Proposition 10 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 $\lambda \in \left[\underline{\underline{\lambda}}(q, c_1, c_2, \eta), \overline{\overline{\lambda}}(q, c_1, c_2, \eta)\right]$, 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

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

Proof: 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:

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

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

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:

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

This strategy will be dynamically consistent as long as:

$$1 - c_1 - c_2 + q\eta - qc_2\eta\lambda \geq -c_1 - (1-q)\eta\lambda + (1-q)c_2\eta \tag{2.26}$$

$$(1-q) - c_1 - (1-q)c_2 - (1-q)q\eta(\lambda - 1)(1 + c_2) \geq \tag{2.27}$$

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

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 (2.26)-(2.27) are satisfied, while if $q = 1$, 2.27 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 2.26 is implied by 2.27 and consequently the strategy will be dynamic consistent as long as inequality 2.27 is satisfied or equivalently as long as

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

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

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

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

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

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

This strategy is dynamic consistent only if:

$$1 - c_1 - c_2 + (1 - \beta(\theta_H, \theta_H)(1-q))\eta - (1 - \beta(\theta_H, \theta_H)(1-q))c_2\eta\lambda =$$
$$- c_1 - \beta(\theta_H, \theta_H)(1-q)\eta\lambda + \beta(\theta_H, \theta_H)(1-q)c_2\eta$$

and

$$(1-q)\beta(\theta_H, \theta_H) - c_1 - c_2(1-q)\beta(\theta_H, \theta_H) -$$
$$- \beta(\theta_H, \theta_H)(1-q)(1 - \beta(\theta_H, \theta_H)(1-q))\eta(\lambda - 1) -$$
$$- (1 - \beta(\theta_H, \theta_H)(1-q))\beta(\theta_H, \theta_H)(1-q)\eta(\lambda - 1)c_2 >$$
$$> \eta(c_1 + \beta(\theta_H, \theta_H)(1-q)c_2 - \beta(\theta_H, \theta_H)(1-q)\lambda)$$

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 $\left(0,0,\hat{\beta}_{LH}^{Tr}\right)$. Note that the total utility associated with this strategy will be 0 and observe that the utility associated with strategy $\left(1,\hat{\beta}_{HL}^{Tr},1\right)$ is greater than the one associated with $\left(0,0,\hat{\beta}_{LH}^{Tr}\right)$ 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:

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

and dynamic consistency would require:[38]

$$\begin{aligned}(1-q)-c_1-(1-q)\,c_2+(1-\alpha\,(1-q))\,(1-q)\,\eta-\alpha\,(1-q)\,q\eta\lambda-\\ -(1-\alpha)\,(c_1+(1-q)\,c_2)\,\eta\lambda-\alpha q\,(1-q)\,c_2\eta\lambda+\alpha\,(1-q)\,qc_2\eta=\\ =-\alpha\,(1-q)\,\eta\lambda+\alpha c_1\eta+\alpha\,(1-q)\,c_2\eta\end{aligned}$$

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

or equivalently:

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

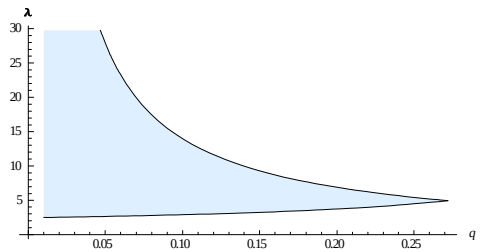
Observe that the utility associated with this strategy will always be lower than 0. The previous analysis implies that as long as $\lambda \leq 1 + \frac{(1-q)(1-c_2)-c_1}{\eta(1-q)q(1+c_2)}$ and

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

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

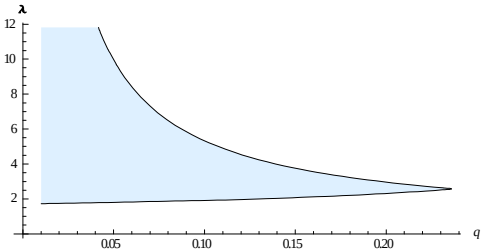
As in the previous section, loss aversion plays a double role. On the one hand, it discourages B from undertaking in the project because this decision exposes him to some risk. To be more precise, suppose B is certain that the project is high-quality and assume further that he always chooses *Stay* whenever 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. Figure 5 represents the set of parameters for which this is indeed the case.



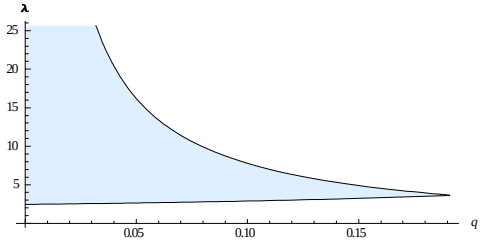
$c_1 = 0.3,\ c_2 = 0.2,\ \eta = 0.3,$

$p_L = 0.15$



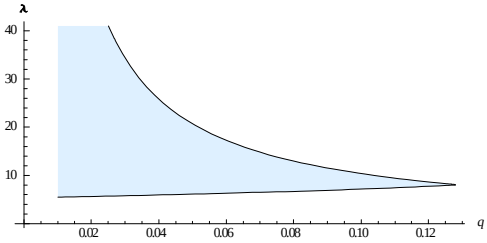
$c_1 = 0.3,\ c_2 = 0.2,\ \eta = 0.9,$

$p_L = 0.15$



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

$p_L = 0.15$



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

$p_L = 0.15$

Figure 5: Set of Parameters for which a Fully Informative Equilibrium Exists.

Proposition 11 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 λ falls in this range 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 10. 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:

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

■

2.4.3 Communication and Reference Dependence in Other 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.[39] 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

[39]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

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.[40]

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 $(1 - p_L)$;[41] 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 previous analysis will hold if the loss aversion of the sender is not too high.[42]

possibly different from $\frac{1}{2}$ and with the additional assumption that the Sender is not informed about the true quality of the project.

[40] Details are available upon request.

[41] 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.

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

2.5 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, 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 2.4.2. 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 2.2 in the obvious way.

Now, we can reinterpret Proposition 9 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.[43]

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.

[43] Given the results in Section 2.4.2, we could include among the set of contingecies the decision of B whether to liquidate the project or not and we would not modify the main insight of this section.

Proposition 12 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[44] 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:

$$\beta((\theta_L, \theta_H), (w, z)) = \begin{cases} 1 & \text{if } \lambda < \frac{(p_H(1+w)+(1-p_H)z)(1+\eta)}{C\eta} - \frac{1}{\eta} \\ x \in [0,1] & \text{if } \lambda = \frac{(p_H(1+w)+(1-p_H)z)(1+\eta)}{C\eta} - \frac{1}{\eta} \\ 0 & \text{if } \lambda > \frac{(p_H(1+w)+(1-p_H)z)(1+\eta)}{C\eta} - \frac{1}{\eta} \end{cases}, \tag{2.28}$$

$k_f^H(\lambda) = 0$ 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: 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 2.4.2 to

[44] 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.

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 2.28.

Suppose instead that $\lambda > \bar{\lambda}(p_H, c_1, c_2, \eta)$. 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 $\lambda > \bar{\lambda}(p_H, c_1, c_2, \eta)$; 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 $k_f^H > 0$.[45] Therefore, for any $\lambda > \bar{\lambda}(p_H, c_1, c_2, \eta)$, the optimal monetary transfers is given by $\left(0, k_f^*\right)$, where:

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

[45] 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.

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 2.4.2. ∎

Proposition 12 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 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) implies $\underline{\lambda}(p_H, c_1, c_2, \eta) < \bar{\lambda}(p_H, c_1, c_2, \eta)$ and Assumption 7(ii) simplifies the characterization 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 2.4.2. 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 12.

Proposition 13 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[46] 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 14 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 $\lambda \in \left[\hat{\lambda}, \underline{\lambda}(p_H, c_1, c_2, \eta)\right]$ 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

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

[46] 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.

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: 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

$$\begin{aligned} p_L (1 + k_s) + (1 - p_L) k_f - C - \\ - p_H (1 - p_L)(1 + k_s - k_f) \eta\lambda + (1 - p_H) p_L (1 + k_s - k_f) \eta \geq \\ \geq -c_1 + c_2\eta - p_H (1 + k_s) \eta\lambda - (1 - p_H) k_f \eta\lambda \end{aligned}$$

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 8 and 9. ∎

Proposition 14 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.

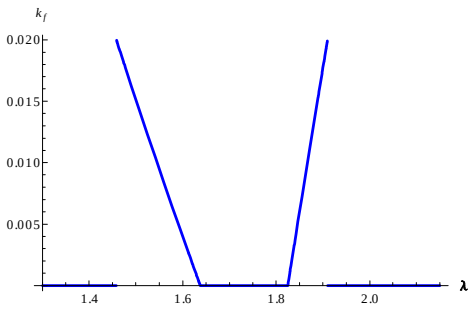


Figure 6: Monetary Transfers as a Function of λ.

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 $\left[\underline{\lambda}\left(p_H, c_1, c_2, \eta\right), \bar{\lambda}\left(p_H, c_1, c_2, \eta\right)\right]$ and increase again (and keep doing so) if $\lambda > \bar{\lambda}\left(p_H, c_1, c_2, \eta\right)$. Figure 6 represents this pattern for a given set of parameters.

2.6 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.[47],[48] 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 reference points determined endogenously on contractual design. As shown in Section 2.5, 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.[49]

[47] 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.

[48] In Grillo (2011a) we considered two senders, but we make the extreme assumption that the "types" of the two senders are independent.

[49] A first step in this direction is provided by de Meza and Webb (2007).

Chapter 3

Reference Dependence and Electoral Speeches

3.1 Electoral Competition and Reference Dependence

"We must not promise what we ought not, lest
we be called on to perform what we cannot."
Abraham Lincoln

The effect of political debates on the outcome of elections has received significant attention in the political science literature; to the very least, these events polarize media's attention and capture the interest of voters: on October 15th 2008, more than 55 millions Americans watched the last debate between the two presidential candidates; furthermore, in the following days, political analysts and communication experts scanned candidates' speeches pointing out incoherences and contradictory statements.

Nevertheless, conventional wisdom concerning electoral promises make two apparently contradictory statements: on the one hand, several people claim that electoral promises lack any credibility since politician would say everything in order to get elected;[1] on the

[1] In contrast with the Lincoln quotes cited above, we can cite the following quote attributed to Napoleon: "If you wish to be a success in the world, promise everything, deliver nothing."

other hand, it is often suggested that excessive electoral promises may create problems for a politician's prospects once voters realize that these promises cannot be delivered.[2]

In this chapter, we present a simple model of communication that is compatible with both these phenomena. In our model, two parties compete for a public office and each party is represented by a candidate who can have two possible valences: high (θ_H) or low (θ_L). The valence of a candidate is his private information, but he can make a public announcement concerning it. The electorate is represented by a continuum of voters who, *coeteris paribus*, prefer high valence candidates to low ones; however, voters differ in ideology and this may lead them to vote based on party's affiliation.

We start showing that in a model in which voters have no reference-dependent attitude, the only equilibrium involves no information transmission: both candidates will claim to have high valence in order to maximize their support in the electorate. This will result in a lack in credibility and voters will ignore electoral promises.

Then, we depart from standard literature by assuming that voters have reference dependent utility a là Köszegi and Rabin. Therefore, we assume that voters evaluate their future prospects with respect to an endogenously determined reference point: whenever the utility experienced by a voter exceeds (falls short of) the reference utility he expects to get, he incurs a psychological gain (loss). We further assume that voters are loss averse, namely that they dislike losses more than what they like gains of the same size.

Under the assumption of reference-dependent utilities, we show that a fully informative equilibrium may be possible. The mechanism behind this equilibrium is as follows: candidates' announcements, if credible, modify the reference point of voters concerning the utility the candidate will deliver; in particular, if a voter biased in favor of party j were to find out that the candidate of party j lied pretending to be high valence while he is not, he may decide to support the other party in order to avoid the loss associated with electing a can-

[2]Indeed, anectodctical evidence suggests the idea that electoral campaign leaders often try to lower the expectations concerning the performance of their candidate even though, a priori, this may be hard to reconcile with standard models of electoral competition (see for example, newspaper articles on George W. Bush's campaign in 2000 and 2004 elections).

didate worse than anticipated. This mechanism is what makes candidates' announcements credible in the first place: candidates will announce their valence truthfully to avoid the risk of losing the support of voters, were a lie be detected. Intuitively, in the model with reference dependence utility, candidates' announcements can modify voters' reference points. Thus, when a low valence candidate has to decide between lying or announcing his valence sincerely, he faces a trade-off. If he is sincere, he loses part of the electorate, but he also ensures a stable support for the remaining of the electoral competition; if he lies, he may increase his electoral support if the lie goes undetected, but he may also decrease it if the lie is found out. Whenever the loss from lying is high enough, the candidate will prefer behaving sincerely. In our model voters can be divided in three different groups depending on their ranking between a low valence candidate of their preferred party and a high valence one from the opposing party. Voters with high ideological bias will prefer the former regardless of his truthfulness, while voters with low ideological bias will always vote for the latter. Finally, voters with intermediate bias will vote for the low valence politician of their preferred party if he has been sincere, but they will prefer the other one if they detect a lie. This type of behavior is due to loss aversion: whereas supporting the poorly skilled candidate may be optimal when the voter was planning to do so, loss aversion makes such a decision more costly when he was expecting someone with better skills.

The mechanism described before induces truthtelling only if the probability with which lies are detected exceeds some threshold. Since an increase in loss aversion has the double effect of decreasing the gain from lying and increasing its cost, we can show that this threshold is decreasing in the degree of loss aversion.

This chapter is organized as follows. In the remaining of this Section, we discuss the relevant literature. In Section 2, we introduce the model and the role played by reference dependence. Section 3 contains a simplified model to convey the same intuition behind our model. Section 4 analyzes the general model and describes the mechanism through which

full information transmission can be attained. Section 5 contains some concludin remarks. Section 3.6 collects the proofs omitted from the main text.

3.1.1 Related Literature

In this paper, we analyze a model of communication in which the Senders (politicians) incur no direct cost from lying; in this sense our paper is related to the extensive literature on strategic information transmission pioneered by Crawford and Sobel (1982) and Green and Stokey (2007).[3] We depart from the existing literature assuming that Receivers (voters in our model) have reference-dependent preferences and we show how this assumption can support credible information transmission.

The paper mostly related to ours is Koszegi (2006). In this paper, the author looks at the role that communication and anticipatory utilities[4] play in an agency problem; however, while he focuses his attention on situations in which the interests of the two parties are perfectly aligned, we assume that the parties have conflicting interests and we tackle the problem of credible information transmission. Furthermore, in our model there are providers of information who compete against each other.[5]

Charness and Dufwenberg (2006) and Charness and Dufwenberg (2011) provide experimental evidence which shows that the decision to exert effort in a joint project is affected

[3]The literature on cheap talk is extensive and we can give only a partial account of it. Farrell and Gibbons (1989), Battaglini (2002), Krishna and Morgan (2001), Krishna and Morgan (2004) and Aumann and Hart (2002) analyze the cases of multiple receivers, multiple renders or multiple communication rounds. Kartik, Ottaviani, and Squintani (2007) and Kartik (2009) relax the assumption of costless communication by introducing a cost from lying. Goltsman, Horner, Pavlov, and Squintani (2009) and Ivanov (2010) look at the effect that a mediatior can have in relaxing the conflict of interest between the sender and the receiver. For a review of the literature on cheap talk, see Farrell and Rabin (1996) and Krishna and Morgan (2008).

[4]On anticipatory utility see Loewenstein (1987), Loewenstein and Prelec (1992), Caplin and Leahy (2001), Epstein (2008).

[5]Another article that introduces anomalies on the side of receivers is Ottaviani and Squintani (2006). In this paper, the authors show that the introduction of naive receivers (namely, receivers who interpret sender's announcements literally and do not take into account sender's interest in distorting their action) together with sophisticated ones can help disciplining the sender to announce the state truthfully.

by cheap messages that could increase the guilt associated with shirking. Although, the mechanism through which communication matters is different, these papers share with ours the idea of looking at the psychological effects of communication and at their consequences on the behavior of Receivers.

The political science literature has long recognized the role played by expectations management in electoral competitions. In particular, Kimball and Patterson (1997) show that the gap between expectations and real performance of politicians play an important role in determining voters' attitude toward Congress.[6] Richard W. Waterman and Silva (1999) extend this analysis by showing that this expectation gap is important in explaining voters' electoral behavior.[7] On a similar note, a growing literature has documented the role played by expectations in the evaluation of public services.[8]

Although this evidence, to the best of our knowledge, no formal model has described the role that voters' expectations and reference points can play in electoral competitions. A partial exception is represented by Lindstadt and Staton (2010), where the authors build a game theoretic setting in which candidates are involved in expectations' manipulation and characterize the conditions under which downward management of expectations is effective in increasing the electoral prospects of candidates. Although our model share some of their motivation, we characterize the actual channel through which expectations can affect electoral behavior and, by doing this, we are able to endogenize the formation of the reference point during electoral campaigns.

Political science literature has extensively studied the effects of candidates' promises and actual performance on electoral competitions. Some scholars have used a Downsian-Hotelling model[9] to investigate electoral competitions in circumstances where candidates

[6]See also G.R. Boynton and Patterson (1969)

[7]See also Sigelman and Knight (1983).

[8]See, for example, James (2009) and the references therein.

[9]Downs (1957) and Hotelling (1929). For a dynamic version of a Downsian model see Duggan and Fey (2006).

can make credible announcements to the electorate.[10] Another line of research pioneered by Farejohn (1986) addresses the conflict of interest between the candidates and the voters; a common feature of these papers is that electoral promises lack any credibility and voters enact dynamic electoral strategy based on actual performance in order to discipline politicians.[11] In this paper, we will assume that candidates have an incentive to lie and lack any instrument to commit themselves to truthtelling; nevertheless they will be able to credibly convey information thanks to the effect that their announcements may have on the reference points of voters.

Banks (1990) builds a model in which the valence of the candidate is unknown, candidates make announcements, but they incur a cost from delivering an outcome different from what announced. In our paper, we endogenize this cost by showing that loss aversion will induce a positive mass of voters to change their electoral behavior after detecting a lie.

The idea that decision makers evaluate their actions based not only on the final outcome they induce, but also on the comparison between these outcomes and a reference point, has been introduced in the economic literature by Kahneman and Tversky (1979); experimental evidence has confirmed the role played by reference points in determining agents' behavior.[12] An important conceptual issue related with reference dependence preferences is the actual formation of reference points; on the one hand, one can assume that the reference point is determined by the status quo;[13] on the other hand, one could take a more forward-looking approach and assume that the reference point of an agent is determined by his future prospects. This latter approach raises the issue of how to close the loop between optimal behavior and reference point's formation; following Koszegi and Rabin (2006), Koszegi and

[10]The credibility can come from assumptions concerning the preferences of the candidates (e.g. purely rent seeking candidates) or by assuming that candidates possess some sort of commitment device.

[11]See Banks and Sundaram (1998) and Berganza (2000), Duggan (2000) and Schwabe (2011).

[12]See Kahneman, Knetsch, and Thaler (1990) and Kahneman, Knetsch, and Thaler (1991), van Dijk and van Knippenberg (1996), and Fehr, Hart, and Zehnder (2011).

[13]About this approach, see Kahneman and Tversky Tversky and Kahneman (1991) and Sugden Sugden (2003b),

Rabin (2007) and Koszegi and Rabin (2009), we will attain this goal assuming rational expectations: the reference point will be determined assuming that agents behave optimally and the optimality of their behavior will be assessed taking into account the reference point induced by it.[14]

The effect of reference dependence and loss aversion on the transmission of information is also analyzed in a companion paper, Grillo (2011b), where we consider the conflict of interest between a sender and a receiver concerning the participation in a risky project.[15]

Finally, in this paper, as well as in Grillo (2011b), the beliefs of agents at earlier nodes in the game affect their behavior at later nodes. In this sense, both these papers can be related to psychological games defined in Geanakoplos, Pearce, and Stacchetti (1989) and Battigalli and Dufwenberg (2009).[16]

3.2 The Model

Consider a 3-period model of electoral competition in which two parties compete to gain the support of a mass of voters. Periods are indexed with $t = 0, 1, 2$ and parties are labelled with r and ℓ. We will refer to these party as to the right- and left-wing party respectively. Each party is represented by a candidate who can have high or low valence (θ_H and θ_L, respectively). The valence of a candidate can be interpreted as some skill or trait of the candidates or as some political position on a dimension on which voters agree; the relevant aspect is that all voters agree that, *coeteris paribus*, a high valence candidate is better than a low valence one. Candidates' party affiliation is publicly observable, while their valence is not. Thus candidates' type are pairs $(i, \theta_k) \in \{\ell, r\} \times \{\theta_L, \theta_H\}$. We assume that types are

[14] For a different approach, see Shalev (2000).

[15] Another paper that looks at the effect of reference points in a contract theory setting is Hart and Moore (2008); for experimental evidence see Fehr, Hart, and Zehnder (2011).

[16] See also Rabin (1993) and Battigalli and Dufwenberg (2007).

chosen according to the following distribution:

	θ_H	θ_L
θ_H	q^2	$q(1-q)$
θ_L	$q(1-q)$	$(1-q)^2$

Thus, q is the *ex-ante* probability with which the candidate of each party has high valence. After their valence has been determined, candidates can make a public statement trying to modify the belief of voters about their type. We assume that communication is costless and that candidates do not incur any cost from lying.

For any n-dimensional vector $(x_1, x_2, ..., x_n)$, we denote with $\sum_{i=1}^{n} p_i [x_i]$ the finite lottery that delivers outcome x_i with probability p_i and with δ_{x_i} the degenerate lottery that delivers outcome x_i with certainty.

The electorate is made by a unit mass of voters who differ in their ideological bias. The bias of agent i is represented by a parameter $\gamma_i \in [-1, 1]$ and $F(.)$ is the cdf representing the empirical distribution of biases in the electorate; we assume that $F(.)$ is absolutely continuous with a pdf $f(.)$ symmetric around $0 : \forall x \in [0, 1]$, $f(x) = f(-x)$. These assumptions are made to simplify the analysis and none of the insights of the paper hinges on them.[17]

Parameter γ_i represents the political bias of agent i. A voter with bias $\gamma_i > 0$, gets utility 1 (respectively, $\kappa < \frac{1}{2}$) by voting for the high (respectively, low) valence candidate of party r. If this voter supports the high (respectively, low) candidate of the left party, he gets a utility of 1 (respectively, κ) with probability $(1 - \gamma_i)$ and of 0 with complementary probability. Therefore $|\gamma_i|$ measure the extent of the ideological bias in favour of party r. The utility of a voter with bias $\gamma_i < 0$ is defined symmetrically. Thus the utility that a voter with bias γ_i

[17] In particular, we could introduce point masses in the distribution. As we explain below, the absolute value of γ_i represents the bias of voter i in favor of one of the parties. Thus, a particularly relevant case would be the one in which the cdf $F(.)$ has a point mass on 0; this would correspond to an electorate where unbiased voters represent a positive mass of the electorate.

gets from voting for each possible candidate can be described as follows:

$\gamma_i < 0$	θ_H	θ_L
ℓ	1	κ
r	$(1+\gamma_i)[1] - \gamma_i[0]$	$(1+\gamma_i)[\kappa] - \gamma_i[0]$

$\gamma_i \geq 0$	θ_H	θ_L
ℓ	$(1-\gamma)[1] + \gamma_i[0]$	$(1-\gamma_i)[\kappa] + \gamma_i[0]$
r	1	κ

where $p[x] + (1-p)[y]$ denotes the lottery that delivers prize x with probability p and prize y with probability $(1-p)$. The structure of voters' payoffs captures two different assumptions: (i) if we focus on party i, all voters prefer a high valence candidate to a low valence one regardless of their ideological bias (this preference is weak if $\gamma_i \in \{-1, 1\}$), and (ii) a voter biased toward party i regards the candidate of party $j \neq i$ as risky: with some probability, this candidate will deliver the same payoff of the equally skilled candidate of the other party, but with complementary probability, this utility will be lower than any of the utilities associated with the types of party i.[18],[19] Thus, with complete information a voter biased in favor of party i would vote for the candidate of party i whenever the candidate of such party is high-valence or both candidates are low valence. In the remaining case (candidate (i, θ_L) faces candidate (j, θ_H), $i \neq j$), the voter would support his favorite party only if the bias is sufficiently high.

Politicians get a utility of 1 from any voter who votes for them.

[18] We choose to set this utility to 0 for simplicity, but the insight of this paper would still hold for any value $x \in (0, \kappa)$.

[19] Observe that the riskiness associated with the candidate of the opposing party is maximal when $\gamma_i = \frac{1}{2}$. Intuitively, if a voter has very low or very high ideological bias, the comparison between candidates of different parties will be more straightforward and less uncertain(in the first case he will regard both candidates as equal, while in the second case, he will consider the candidate of the opposing party as the worst possible choice).

The distribution with which nature selects candidates' types as well as the empirical distribution $F(.)$ is assumed to be common knowledge among players.

The timing of the model is as follows:

- In period -1, the valence of candidates is selected.
- In period 0, each candidate makes a statement concerning his own valence. These statements are public and are listened by both the voters and the candidate of the opposing party. All this is common knowledge among players.
- In period 1, voters receive a public signal concerning the true valence of each candidate. Before voters receive this signal, low valence candidates can exert a costly effort to distort the signal they send.
- In period 2, voters vote and utilities are realized.

The signals voters receive in period 2 can take two values: $s = 0$ and $s = 1$ and the signal structure is as follows. A high valence candidate always sends message $s = 1$, while a low valence one always sends signal $s = 0$. However, low valence candidates can exert a costly effort to distort signal transmission; in particular, if they exert effort $w \in [0, 1]$, the signal received by voters will be 1 with probability w and 0 with probability $(1 - w)$. We assume that exerting effort w entails a quadratic cost of $\frac{cw^2}{2}$. Throughout the paper, we assume that the cost associated with signal distortion is sufficiently high to prevent the candidate from choosing $w = 1$ even if, by doing so, he would get the support of the entire electorate.

Assumption 8 $c > \frac{1}{2}$.

Figure 7 summarizes the structure of the game from the point of view of each vote.

In this model, the behavior of the candidate of party $i \in \{\ell, r\}$ can be described by behavioral strategy (t_i, w_i), where:

- $t_i : \{\theta_L, \theta_H\} \to \Delta(M)$ is the communication strategy of the candidate of party i. Thus, $t_i(\theta_k)[m]$ is the probability with which a candidate with type (i, θ_k) sends message $m \in M$, where M is the set of available messages. Without loss of generality, we will assume that $M = \{\theta_L, \theta_H\}$.

- $w_i : M^2 \to [0,1]$ is the signal-distortion effort of the low valence candidate of party i.[20] Thus, $w_i\left(m^\ell, m^r\right)$ is the effort level exerted by type (i, θ_L) after that candidates have announced $\left(m^\ell, m^r\right)$.[21,22]

To simplify the discussion, we impose that all players agree on the interpretation of the public announcements $\left(m^\ell, m^r\right)$ and that there is common knowledge of this. This assumption, together with the common prior assumption implies that two voters at the same information set share the same belief concerning the valence of the candidates.

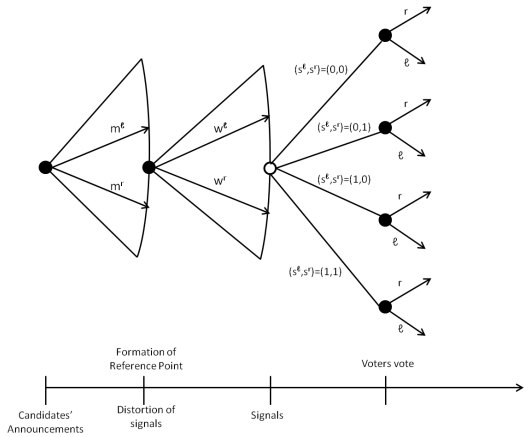


Figure 7: Timing of the Game

[20] Since the high valence candidate always sends signal $s = 1$, he will exert any effort in distorting the signal.

[21] Observe that the effort level of the low type of candidate i may depend on the messages sent by candidates.

[22] We focus on deterministic effort level as opposed to cdf over $[0,1]$ for simplicity. None of the result is affected by this.

We denote the set of histories in the game with $\mathcal{H}$. We denote with $h\left(m^{\ell}, m^{r}\right)$ and $h\left(\left(m^{\ell}, s^{\ell}\right),\left(m^{r}, s^{r}\right)\right)$ be the set of histories compatible with announcements $\left(m^{\ell}, m^{r}\right)$ and announcements-signals pairs $\left(\left(m^{\ell}, s^{\ell}\right),\left(m^{r}, s^{r}\right)\right)$, respectively. Thus

$$\mathcal{I}_{A}=\left\{h\left(m^{\ell}, m^{r}\right): m^{\ell}, m^{r} \in M\right\}$$

will denote the collection of information sets at which voters have listened to announcements $\left(m^{\ell}, m^{r}\right)$ and with

$$\mathcal{I}_{S}=\left\{h\left(\left(m^{\ell}, s^{\ell}\right),\left(m^{r}, s^{r}\right)\right): m^{\ell}, m^{r} \in M \text{ and } s^{\ell}, s^{r} \in\{0,1\}\right\}$$

the collection of information sets at which voters have listened to candidate announcements $\left(m^{\ell}, m^{r}\right)$ and observed signals $\left(s^{\ell}, s^{r}\right)$. We will abuse notation and identify information set $h\left(m^{\ell}, m^{r}\right)$ with $\left(m^{\ell}, m^{r}\right)$ and information set $h\left(\left(m^{\ell}, s^{\ell}\right),\left(m^{r}, s^{r}\right)\right)$ with $\left(\left(m^{\ell}, s^{\ell}\right),\left(m^{r}, s^{r}\right)\right)$.

Although voters take an action only at information sets $\iota_{S} \in \mathcal{I}_{S}$, we assume that the reference point of the agent is formed after listening to candidates' announcements, namely at information sets in $\mathcal{I}_{A}$. Since this reference point will affect the electoral behavior of voters, we will need to model the psychological process experienced by voters at each $\iota_{A} \in \mathcal{I}_{A}$. Finally, we write $\iota_{A} \prec \iota_{S}$ whenever $\iota_{A}=\left(m^{\ell}, m^{r}\right)$ and $\iota_{S}=\left(\left(m^{\ell}, .\right),\left(m^{r}, .\right)\right)$. Thus $\iota_{A} \prec \iota_{S}$ if ι_{S} is compatible with announcements ι_{A}.[23] Let $\mathcal{I}_{S}\left(\iota_{A}\right)=\left\{\iota_{S} \in \mathcal{I}_{S}: \iota_{A} \prec \iota_{S}\right\}$ be the set of feasible information sets given announcements ι_{A}.

The behavior of voters can be described by a function $\rho:[-1,1] \times \mathcal{I}_{S} \rightarrow[0,1]$, where $\rho\left(\gamma_{i}, \iota_{S}\right)$ is the probability with which a voter with bias γ_{i} votes for candidate r at information set ι_{S}. To save on notation, we denote with $\rho\left(. \mid \gamma_{i}, \iota_{A}\right)$ the restriction of $\rho\left(\gamma_{i}, .\right)$ to information sets in $\iota_{S} \in \mathcal{I}_{S}\left(\iota_{A}\right)$. If voters follow strategy ρ, the support for candidates at

[23] Observe that even if ι_{S} is compatible with ι_{A}, voters may assign probability 0 to the possibility of reaching it from ι_{A}.

information set ι_S is given by:

$$\begin{aligned} S_\ell(\iota_S, \rho) &= 1 - \int_{-1}^{1} \rho(\gamma_i, \iota_S)\, dF(\gamma_i) \\ S_r(\iota_S, \rho) &= \int_{-1}^{1} \rho(\gamma_i, \iota_S)\, dF(\gamma_i) \end{aligned}$$

We assume that voters vote sincerely based on their preferences (and beliefs) and that voting is costless.

Suppose that a voter believes that candidates are following behavioral strategy $(t_i, w_i)_{i\in\{\ell,r\}}$. Then, we can apply Bayes rule to compute the probability that the voter assigns to the candidates being high valence at information set $\left(m^\ell, m^r\right) \in \mathcal{I}_A$. This probability is given by:

$$\pi_1^i\left(m^\ell, m^r \mid (t_i, w_i)_{i\in\{\ell,r\}}\right) = \frac{q \cdot t_i(\theta_H)[m^i]}{q \cdot t_i(\theta_H)[m^i] + (1-q)\, t_i(\theta_L)[m^i]} \quad i \in \{\ell, r\}. \tag{3.1}$$

whenever this probability is well defined.[24]

Similarly, we can denote with $\pi_2^i\left(\iota_S \mid (t_i, w_i)_{i\in\{\ell,r\}}\right)$ the probability that voters assign to the candidate of party i being high valence at information set ι_S.[25] This probability will be given by:

$$\pi_2^i((m^\ell, s^\ell), (m^r, s^r) \mid (t_i, w_i)_{i\in\{\ell,r\}}) = \begin{cases} 0 & \text{if } s_i = 0 \text{ or } s_i = 1,\ \pi_1^i(m^\ell, m^r) = 0 \\ & \text{and } w_i(m^\ell, m^r) = 0. \\ \frac{\pi_1^i(m^\ell, m^r)}{\pi_1^i(m^\ell, m^r) + (1-\pi_1^i(m^\ell, m^r))w_i(m^\ell, m^r)} & \text{otherwise} \end{cases}. \tag{3.2}$$

Observe that at information set ι_S, there may be two cases in which Bayes rule is not well defined. The first one arises if $\pi_1^i\left(m^\ell, m^r\right) = 1$ and $s^i = 0$. In this case we will assume that voters will assign probability 0 to the candidate being high valence; to put it differently,

[24] The actual belief after zero-probability messages is irrelevant. Therefore we will assume that all messages are sent with positive probability.

[25] Observe that this belief can depend on the announcement of the candidate of party j, since the effort level of the low type of party i does.

faced with a discrepancy between the hard information represented by signals and the beliefs determined by previous announcements, a voter will decide to adjust his beliefs based on the hard information. The second case is one in which $s^i = 1$, $\pi_1^i\left(m^\ell, m^r\right) = 0$ and $w_i\left(m^\ell, m^r\right) = 0$. In this case, we will assume that voters will keep assigning probability 0 to the candidate being low valence.[26] We will refer to $\pi\left(\left(t_i, w_i\right)_{i\in\{\ell,r\}}\right) = \left(\pi_1^\ell(.), \pi_2^\ell(.), \pi_1^r(.), \pi_2^r(.)\right)$ as to the probability system associated with $\left(t_i, w_i\right)_{i\in\{\ell,r\}}$.[27] When no confusion arises, we will omit the dependency of these probability on $\left(t_i, w_i\right)_{i\in\{\ell,r\}}$.

Finally, given a profile of behavioral strategies for the candidates $\left(t_i, w_i\right)_{i\in\{\ell,r\}}$, for each information set $\iota_A \in \mathcal{I}_A$ we can define the probability with which voters expect to reach information sets $\iota_S \in \mathcal{I}_S(\iota_A)$; we will denote this probability with $\chi\left(. \mid \iota_A, \left(t_i, w_i\right)_{i\in\{\ell,r\}}\right) \in \Delta\left(\mathcal{I}_S(\iota_A)\right)$. Let $\bar{\iota}_A = \left(\bar{m}^\ell, \bar{m}^r\right) \in \mathcal{I}_A$, then:[28]

$$\chi(\iota_S \mid \bar{\iota}_A, \left(t_i, w_i\right)_{i\in\{\ell,r\}}) = \begin{cases} \left(\left(1-\pi_1^\ell\right) w_\ell + \pi_1^\ell\right)\left(\left(1-\pi_1^r\right) w_r + \pi_1^r\right) & \iota_S = \left(\bar{m}^\ell, 1\right), \left(\bar{m}^r, 1\right) \\ \left(\left(1-\pi_1^\ell\right) w_\ell + \pi_1^\ell\right)\left(1-\pi_1^r\right)\left(1-w_r\right) & \iota_S = \left(\bar{m}^\ell, 1\right), \left(\bar{m}^r, 0\right) \\ \left(1-\pi_1^\ell\right)\left(1-w_\ell\right)\left(\left(1-\pi_1^r\right) w_r + \pi_1^r\right) & \iota_S = \left(\left(\bar{m}^\ell, 0\right), \left(\bar{m}^r, 1\right)\right) \\ \left(1-\pi_1^\ell\right)\left(1-w_\ell\right)\left(1-\pi_1^r\right)\left(1-w_r\right) & \iota_S = \left(\left(\bar{m}^\ell, 0\right), \left(\bar{m}^r, 0\right)\right) \end{cases} \tag{3.3}$$

where π_1^i is given by 3.1.

(3.1), (3.2) and (3.3) can be extended to deal with situations in which voters have probabilistic conjectures about the strategy of candidates in standard ways.

Throughout the paper, we will maintain the assumption that players behave optimally given their beliefs, but we make two different assumptions concerning voters' preferences:

[26] This assumption can be dropped assuming that, absent any effort, the low candidate sends signal $s = 1$ with probability ε and signal $s = 0$ with probability $1 - \varepsilon$, with ε small.

[27] Observe that $\pi_1^i\left(m^\ell, m^r\right)$ does not depend on the message m^j and $\pi_2^i\left(\left(m^\ell, s^\ell\right), \left(m^r, s^r\right)\right)$ does not depend on s^j, where $j \neq i$.

[28] To simplify notation we omit to specify the dependency of $\pi_1^i(.)$ and $w_i(.)$ on $\left(\bar{m}^\ell, \bar{m}^r\right)$.

(i) standard expected utility, and (ii) reference-dependent utility a là Kösgezi and Rabin. In the remaining of this section we provide a short introduction to reference dependence and we show how these concepts apply to our model.

3.2.1 Reference Dependent Utility

Let A be a finite set of outcomes and consider an index $u : A \to \mathbb{R}$. We say that an agent has reference-dependent utility if for any pair of outcomes $a, \alpha \in A$, his utility is given by:

$$v(a \mid \alpha) = u(a) + \mu(u(a) - u(\alpha)),$$

where:

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

Thus, the utility function $v(. \mid .)$ depends on two variables: the *actual outcome* a *and the reference outcome* α. In particular $v(a \mid \alpha)$ is the sum of two components: (i) the material utility $u(.)$, and (ii) the gain/loss component, $\mu(.)$. The gain/loss component captures the idea that agents evaluates the utility associated to outcome a, $u(a)$, based on departures from a reference level of utility, $u(\alpha)$. To be more precise, whenever the utility of outcome a exceeds (respectively, falls short of) the reference utility, $u(\alpha)$, the agent experiences a psychological gain (respectively, loss). In this setting η measures the relative importance of the gain/loss component with respect to the material utility, while λ captures the extent of loss aversion, that is the extent to which voters dislike losses more than what they like same-size gains. We make the following two assumptions on these parameters:

Assumption 9 $\eta \in (0, 1)$ and $\lambda > 1$.

Thus, voters will exhibit a positive degree of loss aversion. The assumption that deviation from the reference utility are evaluated linearly can be relaxed at the cost of an increase in notational complexity.

Reference-dependent utility function can be extended to random outcomes $\widetilde{a} \in \Delta(A)$ given a fixed reference point $\alpha \in A$ in the usual way:

$$\forall \tilde{a} \in \Delta(A),\ v(\widetilde{a} \mid \alpha) = \sum_{a \in A} v(a \mid \alpha)\, \widetilde{a}[a],$$

where $\tilde{a}[a]$ is the probability that $\tilde{a}$ assigns to outcome a. Furthermore, we can also extend reference-dependent utilities to incorporate a random reference point:

$$\tilde{a}, \tilde{\alpha} \in \Delta(A),\ v(\tilde{a} \mid \tilde{\alpha}) = \sum_{\alpha \in A} \sum_{a \in A} v(a \mid \alpha)\, \tilde{a}[a]\, \tilde{\alpha}[\alpha]. \tag{3.4}$$

So far, the reference point has been assumed exogenously. One contribution of Koszegi and Rabin (2006), Koszegi and Rabin (2007) and Koszegi and Rabin (2009) is to endogenize the reference point by assuming rational expectations. To be more precise, consider a static setting in which an agent faces a (finite) choice set D from which he has to select an alternative d. Let $\zeta : D \to A$ be the outcome function, mapping each alternative into a final outcome. We will assume that an agent who chooses $d \in D$, foresees that the outcome will be $\zeta(d)$ and consequently that his reference point and reference utility become $\zeta(d)$ and $u(\zeta(d))$, respectively. Following Koszegi and Rabin (2006), Koszegi and Rabin (2007) and Koszegi and Rabin (2009), we can introduce two definitions of equilibrium:

Definition 5 An Unacclimating Personal Equilibrium (UPE) is an action $d \in D$ such that:

$$v(\zeta(d) \mid \zeta(d)) \geq v(\zeta(d') \mid \zeta(d))\ \ \forall d' \in D$$

Definition 6 A Preferred Personal Equilibrium (PPE) is an action $d \in D$ such that, d is a UPE and:

$$v(\zeta(d) \mid \zeta(d)) \geq v(\zeta(d') \mid \zeta(d'))\ \ \forall d' \in D$$

Therefore, whereas in a UPE deviations are evaluated keeping the reference point fixed, in a PPE the agent takes into account the effect of deviations on the formation of the reference point.[29]

In the model of this paper voters are uncertain about the type of candidates. In particular, the utility that a voter gets from voting for type (i, θ_k) when his bias is γ_i and the reference point is α can be described as follows:

$$v\left(i, \theta_k \mid \alpha; \gamma_i\right) = u\left(i, \theta_k; \gamma_i\right) + \mu\left(u\left(i, \theta_k; \gamma_i\right) - u\left(\alpha\right)\right) \quad i \in \{\ell, r\}, \ k \in \{L, H\},$$

where

$$u\left(i, \theta_k; \gamma_i\right) = \begin{cases} 1 & \text{if } (i, \theta_k) = (r, \theta_H) \\ \kappa & \text{if } (i, \theta_k) = (r, \theta_L) \\ (1 - \gamma_i)[1] + \gamma_i[0] & \text{if } (i, \theta_k) = (\ell, \theta_H) \\ (1 - \gamma_i)[\kappa] + \gamma_i[0] & \text{if } (i, \theta_k) = (\ell, \theta_L) \end{cases}$$

if $\gamma_i \geq 0$ and

$$u\left(i, \theta_k; \gamma_i\right) = \begin{cases} (1 - \gamma_i)[1] + \gamma_i[0] & \text{if } (i, \theta_k) = (r, \theta_H) \\ (1 - \gamma_i)[\kappa] + \gamma_i[0] & \text{if } (i, \theta_k) = (r, \theta_L) \\ 1 & \text{if } (i, \theta_k) = (\ell, \theta_H) \\ \kappa & \text{if } (i, \theta_k) = (\ell, \theta_L) \end{cases}$$

if $\gamma_i < 0$. Observe that if $\eta = 0$, voters would care about material utility only and they would behave like expected utility maximizers with vNM utility index given by $u(.)$. Since terminal

[29] Koszegi (2010) provide conditions for the existence of PPEs.

nodes are associated with utilities in a unique way, we will deal directly with reference utilities without specifying the reference points associated with them.

At information set $\iota_S \in \mathcal{I}_S$, the belief of a voter concerning candidates' valence can be represented by π^ℓ and π^r, where π^i is the probability with which the candidate of party i has high valence. Let $\tilde{u}\left(x;\gamma_i,\pi^\ell,\pi^r\right)$ denote the lottery over utilities induced by a voter who votes for the candidate of party r (ℓ) with probability x $(1-x)$, has bias γ_i and has beliefs $\left(\pi^\ell,\pi^r\right)$. If $\gamma_i \geq 0$, this lottery will be given by:

$$\tilde{u}\left(x;\gamma_i,\pi^\ell,\pi^r\right)[y] = \begin{cases} x\pi^r + (1-x)\,\pi^\ell\,(1-\gamma_i) & \text{if } y = 1 \\ x\,(1-\pi^r) + (1-x)\left(1-\pi^\ell\right)(1-\gamma_i) & \text{if } y = \kappa \\ (1-x)\,\gamma_i & \text{if } y = 0 \end{cases}$$

The expression for voters with negative bias is analogous.

In this paper, we account for the fact that candidates' announcement, if credible, can affect the reference point of voters and may modify their electoral behavior. This point deserves some further discussion. Suppose that after announcements $\iota_A = \left(m^\ell, m^r\right)$ voters formulate a strategy concerning their electoral behavior for each information set $\iota_S \in \mathcal{I}_S\left(\iota_A\right)$. We will assume that this thought process determines the reference utility of voters and that, once this reference utility is determined, voter will take it as given. To put it differently, we will assume that the reference utility of voters is determined at information sets $\mathcal{I}_A$ and may affect voters' electoral behavior at information sets $\mathcal{I}_S$. This timing captures the idea that, if the announcements of candidates modify the beliefs of voters, they will also modify their future prospects and, consequently, their reference utility; the assumption that the reference point is determined by announcements only is made for simplicity: the main insight of the paper would still hold if we were to assume that the reference point is determined by a

weighted sum of the information received through the announcements of the one received through signals.

The formation of reference points at information sets $\iota_A \in \mathcal{I}_A$ raises two type of issues: the consistency of reference points with the actual behavior of voters and the possibility of multiple consistent reference points. To understand the first point, note that, in our model, at information set $\iota_A \in \mathcal{I}_A$, a voter (i) modifies his belief concerning candidates' valences and (ii) formulates a plan that describes his behavior for any information set he may reach in period 2. The plan formulated by a voter with bias γ_i can be summarized by a continuation strategy $\hat{\rho}(. \mid \gamma_i, \iota_A) : \mathcal{I}_S(\iota_A) \to [0,1]$, where $\hat{\rho}(\iota_S \mid \gamma_i, \iota_A)$ represents the probability with which the voter plans to vote for party r at information ι_S given that he is at information set is ι_A. Since the beliefs over information sets in $\mathcal{I}_S(\iota_A)$ are represented by $\hat{\chi}(.) \in \Delta(\mathcal{I}_S(\iota_A))$ and the beliefs about candidates' types at each information set $\iota_S \in \mathcal{I}_S(\iota_A)$ are represented by a pair of functions $\hat{\pi}_2^i : \mathcal{I}_S(\iota_A) \to [0,1]$ $(i \in \{\ell, r\})$, the reference utility at information set ι_A will be a lottery that assigns probability

$$\sum_{\iota_S} \hat{\chi}(\iota_S) \cdot \tilde{u}\left(\hat{\rho}(\iota_S \mid \gamma_i, \iota_A); \gamma_i, \hat{\pi}_2^{\ell}(\iota_S), \hat{\pi}_2^{r}(\iota_S)\right)[y]$$

to every outcome $y \in \mathbb{R}$.

Now suppose that a voter is certain that candidates are following behavioral strategy $(t_i, w_i)_{i\in\{\ell,r\}}$. Thus, if he plans to follow strategy $\rho(\gamma_i, .)$, his reference utility at information set ι_A will be given by $\tilde{u}\left(\rho; \gamma_i, \iota_A, (t_i, w_i)_{i\in\{\ell,r\}}\right)$, where for every $y \in \mathbb{R}$:[30]

$$\tilde{u}\left(\rho; \gamma_i, \iota_A, (t_i, w_i)_{i\in\{\ell,r\}}\right)[y] = \\ = \sum_{\iota_S} \chi(\iota_S \mid \iota_A) \cdot \tilde{u}\left(\rho(\iota_S \mid \gamma_i, \iota_A); \gamma_i, \pi_2^{\ell}(\iota_S \mid \iota_A), \pi_2^{r}(\iota_S \mid \iota_A)\right)[y]$$

[30]In what follows, we omit to specify the dependency of $\pi_2^i\left(\iota_S \mid \iota_A, (t_i, w_i)_{i\in\{\ell,r\}}\right)$ and $\chi\left(\iota_S \mid \iota_A, (t_i, w_i)_{i\in\{\ell,r\}}\right)$ on $(t_i, w_i)_{i\in\{\ell,r\}}$

We will say that a strategy ρ is *dynamic consistent* at information set ι_A for voter γ_i given $(t_i, w_i)_{i\in\{\ell,r\}}$ if, for each information set $\iota_S \in \mathcal{I}_S(\iota_A)$, $\rho(\iota_S \mid \gamma_i; \iota_A)$ is optimal given reference utility $\tilde{u}\left(\rho; \gamma_i, \iota_A, (t_i, w_i)_{i\in\{\ell,r\}}\right)$. Thus, dynamic consistency requires $\rho(\gamma_i, .)$ to specify a behavior that is optimal under the assumption that the reference utility is determined according to such behavior. In this sense, dynamic consistency imposes an constraint in the spirit of UPE and evaluates deviations keeping the reference utility fixed.

To define this concept formally, we need to introduce some additional notation. Let $\tilde{u}_1$ and $\tilde{u}_2$ be two finite lotteries over $\mathbb{R}$. Then:

$$E\tilde{u}_1 := \sum_x \tilde{u}_1[x] \cdot x$$

and:

$$\mu(\tilde{u}_1 - \tilde{u}_2) := \sum_y \sum_x \mu(\tilde{u}_1[x] \cdot x - \tilde{u}_2[y] \cdot y)$$

where for any $r \in \mathbb{R}$, $\mu(r)$ is defined by (2.5). We are now ready to provide the following definition:

Definition 7 Strategy $\rho(.,.)$ is dynamic consistent at information set ι_A for bias γ_i given $(t_i, w_i)_{i\in\{\ell,r\}}$, if $\forall \iota_S \in I_S(\iota_A)$ and $\forall y \in [0,1]$

$$\begin{aligned} &E\tilde{u}\left(\rho(\gamma_i, \iota_S); \gamma_i, \pi_2^\ell(\iota_S), \pi_2^r(\iota_S)\right) + \\ &\quad + \mu\left(\tilde{u}\left(\rho(\gamma_i, \iota_S); \gamma_i, \pi_2^\ell(\iota_S), \pi_2^r(\iota_S)\right) - \tilde{u}\left(\rho; \gamma_i, \iota_A, (t_i, w_i)_{i\in\{\ell,r\}}\right)\right) \geq \\ &E\tilde{u}\left(y; \gamma_i, \pi_2^\ell(\iota_S), \pi_2^r(\iota_S)\right) + \mu\left(\tilde{u}\left(y; \gamma_i, \pi_2^\ell(\iota_S), \pi_2^r(\iota_S)\right) - \tilde{u}\left(\rho; \gamma_i, \iota_A, (t_i, w_i)_{i\in\{\ell,r\}}\right)\right) \end{aligned}$$

Strategy $\rho(.,.)$ is dynamic consistent given $(t_i, w_i)_{i\in\{\ell,r\}}$ if it is dynamic consistent at each information set $\iota_A \in I_A$ for every γ_i given $(t_i, w_i)_{i\in\{\ell,r\}}$.

Although dynamic consistency rules out incoherent strategies, it does not rule out the existence of multiple dynamic consistent strategy. Consider the following example:

Example 4 Suppose $\kappa = \frac{2}{5}$ and consider a voter with bias $\gamma_i = \frac{1}{2}$. Let $\pi_1^{\ell}\left(m^{\ell}, m^{r}\right) = 1$, $\pi_2^{r}\left(m^{\ell}, m^{r}\right) = 0$. Thus $\chi\left(\iota_S \mid \iota_A\right) > 0$ only if $s^{\ell} = 1$. Furthermore, assume that the voter behaves in the same way after information sets $\left(\left(m^{\ell}, 1\right), \left(m^{r}, 0\right)\right)$ and $\left(\left(m^{\ell}, 1\right), \left(m^{r}, 1\right)\right)$.[31] In this case, the voter assign probability 1 to the candidate of the left party sending signal $s^{\ell} = 1$ and his reference utility can be determined taking into account the behavior at these nodes only.[32] Assume that the voter thinks of voting for the candidate of the right party. In this case his reference utility would be $\delta_{\frac{2}{5}}$. This strategy will be dynamically consistent as long as:

$$\frac{2}{5} > \frac{1}{2} + \frac{1}{2}\left(1 - \frac{2}{5}\right)\eta + \frac{1}{2} \cdot \left(0 - \frac{2}{5}\right)\eta\lambda$$

or equivalently:

$$\lambda \geq \frac{3}{2} + \frac{1}{2\eta}$$

Suppose instead that the voter is planning to support the candidate of the left party. In this case his reference utility will be 1 with probability $\frac{1}{2}$ and 0 with probability $\frac{1}{2}$. This will be dynamic consistent as long as:

$$\frac{1}{2} - \frac{1}{4}\eta\left(\lambda - 1\right) \geq \frac{2}{5} + \frac{1}{2}\left(\frac{2}{5} - 1\right)\eta\lambda + \frac{1}{2}\left(\frac{2}{5} - 0\right)\eta$$

which is always satisfied. Then as long as $\lambda \geq \frac{3}{2} + \frac{1}{2\eta}$, we have at least two dynamic consistent strategies.[33]

Whenever this multiplicity arises, we will assume that the voter selects the dynamic consistent strategy that maximizes his utility at information set ι_A. Although, we will model the agent as consciously choosing the optimal dynamic consistent strategy, we could equiv-

[31] Our assumtion about beliefs imply that voters will have the same beliefs at each of these information nodes. Furthermore, one can show that this type of behavior will indeed be the "optimal" one.

[32] Of course, the actual equilibrium strategy will prescribe a behavior for the other nodes as well, but this behavior will be irrelevant in determining the reference utility.

[33] Actually, there is also a third mixed dynamically consistent strategy, but its characterization is irrelevant at the moment.

alently interpret this choice as an unconscious behavior. Furthermore, the assumption that the voter always selects the "optimal" reference point can be relaxed with the assumption that the optimal reference point is chosen with some positive probability without affecting the main insight of the paper.[34]

Example 5 (continued) Consider the previous example and suppose that $\lambda \geq \frac{3}{2} + \frac{1}{2\eta}$. Thus, we have already shown that we have two dynamic consistent strategies. In particular, the utility associated with the second dynamic consistent strategy is maximized at $\lambda = \frac{3}{2} + \frac{1}{2\eta}$ and in this case it is equal to $\frac{3}{8} - \frac{1}{8}\eta < \frac{2}{5}$. Thus, the former strategy provides a higher total utility than the second and we conclude that the voter will prefer following it.[35]

We are now ready to define the equilibrium concept we will use in the analysis of the model with reference dependent utility.

Definition 8 An equilibrium of the game with reference dependent utility is a profile of behavioral strategies $\left(\left(t_i^*, w_i^*\right)_{i\in\{\ell,r\}}, \rho^*\right)$ and a belief system $\pi\left(\left(t_i^*, w_i^*\right)_{i\in\{\ell,r\}}\right)$ such that:
(i) $\rho^*(.,.)$ is dynamically consistent given $\left(t_i^*, w_i^*\right)_{i\in\{\ell,r\}}$ and for any other dynamic consistent strategy $\rho(.,.)$ given $\left(t_i^*, w_i^*\right)_{i\in\{\ell,r\}}$, any information set $\iota_A \in I_A$ and every γ_i :[36]

$$\sum_{\iota_S} E\tilde{u}\left(\rho^*\left(\gamma_i, \iota_S\right); \gamma_i, \pi_2^\ell\left(\iota_S\right), \pi_2^r\left(\iota_S\right)\right) - \sum_{\iota_S} E\tilde{u}\left(\rho\left(\gamma_i, \iota_S\right); \gamma_i, \pi_2^\ell\left(\iota_S\right), \pi_2^r\left(\iota_S\right)\right) \geq$$

$$\geq \sum_{\iota_S} \mu\left(u\left(\rho\left(\gamma_i, \iota_S\right); \gamma_i, \pi_2^\ell\left(\iota_S\right), \pi_2^r\left(\iota_S\right)\right) - \hat{u}\left(\rho, ; \gamma_i, \iota_A\right)\right) \chi\left(\iota_S \mid \iota_A\right) -$$

$$- \sum_{\iota_S} \mu\left(u\left(\rho^*\left(\gamma_i, \iota_S\right); \gamma_i, \pi_2^\ell\left(\iota_S\right), \pi_2^r\left(\iota_S\right)\right) - \hat{u}\left(\rho^*; \gamma_i, \iota_A\right)\right) \cdot \chi\left(\iota_S \mid \iota_A\right)$$

[34] For a model in which agents choose their future expectations optimally, see Brunnermeier and Parker Brunnermeier and Parker (2005).

[35] One can further show that this is the optimal strategy among the set of all dynamic consistent strategies.

[36] To simplify notation, in the following expression, we omit the dependency on $\left(t_i^*, w_i^*\right)_{i\in\{\ell,r\}}$.

(ii) (t_i^*, w_i^*) maximizes $S_i(\iota_S, \rho^*)$ given $\left(t_j^*, w_j^*\right)$ $j \neq i$,

(iii) $\pi\left((t_i^*, w_i^*)_{i\in\{\ell,r\}}\right)$ is determined by $(t_i^*, w_i^*)_{i\in\{\ell,r\}}$ using (3.1) and (3.2).

Part(i) of Definition 8 capture the idea that ρ^* must be dynamic consistent and must maximize voters' utility among the set of dynamic consistent strategies. In this sense, it captures the optimality criterion of a PPE. We will refer to a dynamic consistent strategy that satisfy condition (i) as to the *optimal dynamic consistent strategy* given $(t_i^*, w_i^*)_{i\in\{\ell,r\}}$.

In the paper, we will focus on two particular class of equilibria: uninformative and fully revealing equilibria:

Definition 9 Let $\left((t_i^*, w_i^*)_{i\in\{\ell,r\}}, \rho^*\right)$ and $\pi\left((t_i^*, w_i^*)_{i\in\{\ell,r\}}\right)$, be an equilibrium. Then:
- the equilibrium is fully informative if for every $i \in \{\ell, r\}$ and for every m, $t_i^*(\theta_k)[m] > 0$ implies $t_i^*(\theta_s)[m] = 0$ $s, k \in \{L, H\}$, $s \neq k$.
- the equilibrium is uninformative if for every $i \in \{\ell, r\}$ and for every m, $t_i^*(\theta_L)[m] = t_i^*(\theta_H)[m]$.

In the remaining of the paper, we will denote the communication strategy of candidate i in a fully informative equilibrium with t_i^{Tr} and the one in a uninformative equilibrium with t_i^{Un}.

3.3 A Simplified Model

Before analyzing the model we described in Section 3.2, we consider a simpler version of the model that will help understanding the mechanism behind our result. To this goal, we modify the model as follows:

1. The electorate is made by three voters: LB,U and RB. The expected utility each voter gets by voting for the different candidates is summarized in the following tables:

LB	θ_H	θ_L
ℓ	1	$\frac{2}{5}$
r	$\frac{1}{2}[1]+\frac{1}{2}[0]$	$\frac{1}{2}\left[\frac{2}{5}\right]+\frac{1}{2}[0]$

,

U	θ_H	θ_L
ℓ	1	$\frac{2}{5}$
r	1	$\frac{2}{5}$

,

RB	θ_H	θ_L
ℓ	$\frac{1}{2}[1]+\frac{1}{2}[0]$	$\frac{1}{2}\left[\frac{2}{5}\right]+\frac{1}{2}[0]$
r	1	$\frac{2}{5}$

.

Thus LB (respectively, RB) is a voter biased in favor of party ℓ (respectively, r), while U is an unbiased. These payoffs can be mapped in the general model assuming that $\kappa = \frac{2}{5}$, $\gamma_{\text{LB}} = -\frac{1}{2}$, $\gamma_{\text{U}} = 0$ and $\gamma_{\text{RB}} = \frac{1}{2}$.

2. The signal structure is as follows. There are two possible signals: $s = 0$ and $s = 1$. High valence candidates always send signal $s = 1$, while low valence ones will send each signal with probability $\frac{1}{2}$. Furthermore, low valence candidates cannot exert any effort to distort the signal. Thus, regardless of the actual announcement, there exists an exogenous probability (namely, $\frac{1}{2}$) with which low valence is revealed.

3. $q = \frac{1}{2}$, so that all types of candidates are equally likely.

4. $x \in \left(0, \frac{1}{2}\right)$ represents the empirical frequency of left- and right-biased voters in the population.[37] The frequency of unbiased voters is thus equal to $(1 - 2x)$. Thus candi-

[37] Alternatively, differences in the relative importance of voters may be justified assuming that the support of the biased voters is more important than that of unbiased voters ($x > \frac{1}{3}$):

dates get a utility equal to x if they obtain the support of each group of biased voters and equal to $1-2x$ if they obtain the support of the unbiased voters.

To simplify the discussion we will also assume that, whenever indifferent between two candidates, unbiased voters randomize with equal probability, while biased ones vote based on their ideological bias.[38]

We begin considering the benchmark case in which $\eta = 0$. This corresponds to a situation in which voters care about material utility only. The following proposition shows that, under this assumption, candidates' announcements will lack any credibility and that the behavior of voters will not depend on these announcements.

Proposition 15 If $\eta = 0$, in every equilibrium candidates announcements do not convey any information concerning their types.

Proof: Let $\pi^i(m^i, s^i)$ denote the probability voters assign to the candidate of party i having high valence after listening to candidates announcements m^i and observing signal s^i.[39] Our assumptions on voters' updating imply that for any pair of message m^i, $\pi^i(m^i, 0) = 0$. At information set $((m^i, s^i), (m^j, s^j))$, U would vote for party i (j) if $\pi^i(m^i, s^i) > (<) \pi^j(m^j, s^j)$ and will randomize with equal probability if $\pi^i(m^i, s^i) = \pi^j(m^j, s^j)$.

Now consider voter RB. He would vote for the candidate of party r (ℓ) if:

$$\pi^r(m^r, s^r) + \frac{2}{5}(1 - \pi^r(m^r, s^r)) > (<) \frac{\pi^\ell(m^\ell, s^\ell)}{2} + \frac{1}{5}\left(1 - \pi^\ell\left(m^\ell, s^\ell\right)\right)$$

or equivalently:

$$\pi^r(m^r, s^r) > (<) \frac{1}{2}\pi^\ell\left(m^\ell, s^\ell\right) - \frac{1}{3}$$

indeed, biased voters may volunteer for the electoral campaign, make donations in favor or they may simply be more likely to vote.

[38] The assumption that ties are broken in this way is made for simplicty and it does not play any role in the general model. Observe that we fix the tie-breaking rule so that voters will punish candidates for lying only if they have a strict incentive to do so.

[39] Since candidates moves simultaneously and, in this section, do not take any further action after that announcements have been made, this belief will not depend on message m^j or on signal s^j.

If the previous inequality hold with equality, RB will vote for the candidate of party r. The analysis for voter LB is identical and omitted. Now suppose, by contradiction with our statement, that there exists an equilibrium in which the announcement of one of the candidates convey some information concerning candidate's valence. Then we can find a party i and two messages, m_L^i and m_H^i, such that type θ_L of party i sends message m_L^i more often than message m_H^i. Let $\underline{m}$ be the message associated with the highest probability of being low type and let k be the party of the candidate who send this message. Then there must be another message $\bar{m}$ sent by the candidate of party k, such that $\pi^k(\underline{m}, 0) = \pi^k(\bar{m}, 0) = 0$ and $\pi\left(\underline{m}^k, 1\right) < \frac{2}{3} < \pi\left(\bar{m}^k, 1\right)$. Furthermore, by construction there must be a message m^j sent by the candidates of party $j \neq k$ is such that $\pi\left(\underline{m}^k, 1\right) \leq \pi(m^j, 1) < \pi\left(\bar{m}^k, 1\right)$. We conclude that by sending message $\bar{m}^k$ instead of $\underline{m}^k$, the low type would increase his payoff (since he would get the support of the unbiased voter whenever the types of the other candidates are sending message m^j). This contradicts the assumption that the strategy followed by the types of party k behaves optimally. ∎

The intuition behind this result is as follow: without reference dependence, candidates announcements affect only the beliefs of voters. Since a candidate is more likely to get the support of voters if they believe he is a high-valence candidate and since lies are costless, candidates will always send the message that maximizes the posterior probability of being high valence. As a result, their announcements will lack any credibility and voters will ignore them. The introduction of reference dependence and loss aversion modifies this result through the effect announcements have on the formation of the reference utility. Indeed, the next proposition shows that in this case there exists a fully revealing equilibrium in which candidates announce their type truthfully and voters believe in these announcements.

Proposition 16 Suppose that $\eta > 0$. Then, if $x \geq \frac{1}{3}$, there exists a threshold level $\bar{\lambda}(\eta)$ such that if $\lambda > \bar{\lambda}(\eta)$, a fully revealing equilibrium exists.

Proof: Suppose that candidates announced their types truthfully in period 0. Then, we can assume that the set of feasible messages is given by $M = \{\theta_L, \theta_H\}$ and that following message $m^i = \theta_k$, all voters will assign probability 1 to the candidate of party i being type θ_k.

Consider voter U, first. In a truthful equilibrium the reference utility associated with the unique dynamically consistent strategy would be a degenerate measure on 1 if at least one candidate sent message $m = \theta_H$ and a degenerate measure on $\frac{2}{5}$ otherwise. Indeed suppose that the candidate of party i announced $m^i = \theta_H$. Then U believes that by voting for him he can get a utility of 1. If the candidate of party j also announced $m^j = \theta_H$, the voter will assign probability one to his utility being 1 regardless of his actual electoral behavior. If instead $m^j = \theta_L$ and the voter was planning to vote for the candidate of party j, his reference utility would be $\frac{2}{5}$ and voting for the candidate of party i would lead to a higher utility. Given these reference utility, the electoral behavior of voters at information set ι_S can be described as follows:

- vote for the candidate of party r if

$$\iota_S \in \{((\theta_H, 0), (\theta_H, 1)), ((\theta_L, 0), (\theta_H, 1)), ((\theta_L, 1), (\theta_H, 1))\}$$

- vote for the candidate of party ℓ if

$$\iota_S \in \{((\theta_H, 1), (\theta_L, 1)), ((\theta_H, 1), (\theta_L, 0)), ((\theta_H, 1), (\theta_H, 0))\}$$

- randomize with equal probability between the two candidates at all remaining information sets.[40]

Now, consider voter RB. It is immediate to show that the unique dynamic consistent strategy

[40]Observe that we assume that the voter U does not punish the candidate who lied and randomizes with equal probability whenever the two candidates are associated with the same utility. If we were to break indifferences in a way that is less favourable way for a liar, we would reinforce the mechanism that leads to truthful information transmission.

must prescribe to vote for the candidate of party r following announcements (θ_L, θ_L) and (θ_L, θ_H), regardless of the actual signal received and of the truthfulness of party r's candidate. In the first case, the reference utility of the voter would be $\frac{2}{5}$, while in the latter case it would be 1.

Now, suppose candidates sent message (θ_H, θ_H). In this case, the unique dynamic consistent strategy is to vote for the candidate of party r and consequently, the reference utility of the voter would be a degenerate measure on 1. By definition of dynamic consistent strategy, if the signals do not falsify the announcement of candidates (that is both candidates send signal $s = 1$), RB would vote indeed vote for the candidate of party r. Now, suppose that the the signal sent by r were to reveal that he lied (while the signal sent by ℓ is compatible with his initial announcement). Then, if RB were to vote for his preferred party, his utility would be $\frac{2}{5} + \left(\frac{2}{5} - 1\right)\eta\lambda$, while voting for the candidate of the left party would give him a utility equal to $\frac{1}{2} + \frac{1}{2}(1-1)\eta + \frac{1}{2}(0-1)\eta\lambda$. Since the former utility is lower that the latter, we can conclude that in such a situation RB would vote for the candidate of party ℓ. A full characterization of the equilibrium would require us to describe the behavior at other information sets ι_S as well, but this is irrelevant for characterizing the fully informative equilibrium and consequently we will postpone this task until we analyze the general model. Finally, suppose that candidates sent messages (θ_H, θ_L). If RB were planning to vote for the candidate of the right party his reference utility would be a degenerate measure on $\frac{2}{5}$. This would be dynamically consistent as long as:

$$\frac{2}{5} \geq \frac{1}{2} + \frac{1}{2}\left(1 - \frac{2}{5}\right)\eta + \frac{1}{2}\left(0 - \frac{2}{5}\right)\eta\lambda$$

which is satisfied if $\lambda \geq \frac{3}{2} + \frac{1}{2\eta}$. Define:

$$\lambda^*(\eta) := \frac{3}{2} + \frac{1}{2\eta}$$

and observe that $\lambda^*(\eta)$ is decreasing in η. On the other hand, if he were planning to vote for the candidate of party ℓ, his reference utility would be a measure that assign probability $\frac{1}{2}$ to 1 and probability $\frac{1}{2}$ to 0. This would be dynamically consistent as long as:

$$\frac{1}{2} - \frac{1}{4}\eta(\lambda - 1) \geq \frac{2}{5} + \frac{1}{2}\left(\frac{2}{5} - 1\right)\eta\lambda + \frac{1}{2}\left(\frac{2}{5} - 0\right)\eta$$

which is always satisfied. Thus if $\lambda \geq \lambda^*(\eta)$, there will be two dynamic consistent strategies. However the utility associated with the former strategy will always be higher than the one associated with the latter since for any admissable value of λ:

$$\frac{1}{2} - \frac{1}{4}\eta(\lambda - 1) \leq \frac{3}{8} - \frac{1}{8}\eta \leq \frac{2}{5}$$

We conclude that as long as $\lambda \geq \lambda^*(\eta)$, the reference utility associated with the optimal dynamic consistent strategy will be a degenerate measure on $\frac{2}{5}$. Given the symmetric nature of the game, the characterization of LB's electoral behavior is similar to RB's one and omitted.

Now, we will focus our attention on the candidate of party i and verify that, given the voters' behavior we just described and given that the candidate of party $j \neq i$ is announcing his own type truthfully, he will announce his type truthfully as long as $\lambda \geq \lambda^*(\eta)$.

Pick party r (the analysis for party ℓ is identical). Clearly, the high-valence candidate will always want to send message $m^i = \theta_H$ because in this way he will maximize the probability to get the support of the unbiased voter and he will also win the support of RB (while he assigns probability 0 to the possibility of convincing LB to vote for him).

Now consider type θ_L. If he were to reveal his type truthfully and make announcement $m = \theta_L$, he would: (i) get the support of RB with certainty, (ii) get the support of U with probability $\frac{1}{2}$ only if the other candidate also sent message $m = \theta_L$ (given that we are considering a truthful equilibrium, this would happen with probability $\frac{1}{2}$), and (iii) never get the support of LB. Thus the total utility from telling the truth would be: $\frac{1}{2}x + \frac{1}{4}$. If type

θ_L, were to lie and announce to be type θ_H, he would: (i) get the support of RB if the other candidate announced to be a low type (which happens with probability $\frac{1}{2}$) or if the other candidate announced to be a high type and his own lie goes undetected (this happens with probability $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$), (ii), get the support of U with probability 1 if $m^\ell = \theta_L$ and his own lie is not detected (this happens with probability $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$) and with probability $\frac{1}{2}$ if either $m^\ell = \theta_L$ and his own lie is detected or $m^\ell = \theta_H$ and his own lie goes undetected (both these events happen with probability $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$) and (iii) never get the support of LB. Thus the total utility from lying would be $\frac{1}{2} - \frac{1}{4}x$. Thus the candidate of party r will tell the truth as long as $x \geq \frac{1}{3}$. ∎

Proposition 16 implies that a sufficiently high degree of loss aversion together with some assumption concerning the distribution of the electorate may induce full information transmission. Intuitively, candidates' announcements, if credible, modify the reference utility of voters; thus, loss aversion will induce an endogenous and credible punishment against a candidate who lied: to avoid the loss associated with voting for a candidate who is worse than what anticipated, a biased voter may decide to change his electoral behavior and vote for the candidate of his least preferred party. Thus, lying is a risky strategy for the candidate: with some probability his lie will go undetected and he will increase his expected support in the electorate, but with complementary probability he will be caught lying and he will lose the support of biased voters; if the mass of biased voters is high enough, the candidate may prefer announcing his valence truthfully. To put it differently, whereas in the model without reference dependence candidates were competing with a costless instrument (cheap announcements) to get the vote of voters, the introduction of reference dependence adds a cost from lying: were a lie be detected, the candidate would not only lose the support of unbiased voters (who would have not voted for him even if he had announced his type truthfully), but also of the voters biased in his favor (who, instead, would have voted for him, if he had been sincere from the beginning).

Although the previous propositions convey the main intuition of the paper, it leaves open some question: how do ideological biases interact with loss aversion? Which features of the empirical distribution of voters are important to support the fully informative equilibrium? How do these features change with changes in parameters? To answer these questions, we will now analyze the general model described in Section 3.2.

3.4 Analysis of the General Model

3.4.1 Voters with Standard Utility

We start our analysis by considering the model in which the psychological component does not affect the behavior of voters ($\eta = 0$); this can be considered equivalent to the case in which voters are expected utilities maximizers with vNM utility indexes given by $u(i, \theta_k; \gamma_i)$. Under this assumption, it is easy to verify that the optimal electoral behavior of a voter with beliefs $\left(\pi_2^{\ell}(.), \pi_2^{r}(.)\right)$ can be described as follows. For each information set $\iota_S \in \mathcal{I}_S$, the behavior of voters with positive bias ($\gamma_i > 0$) is given by:

$$\rho(\gamma_i, \iota) = \begin{cases} 1 & \pi_2^r(\iota) + (1 - \pi_2^r(\iota))\kappa > \pi_2^{\ell}(\iota)(1 - \gamma_i) + \left(1 - \pi_2^{\ell}(\iota)\right)(1 - \gamma_i)\kappa \\ x & \pi_2^r(\iota) + (1 - \pi_2^r(\iota))\kappa = \pi_2^{\ell}(\iota)(1 - \gamma_i) + \left(1 - \pi_2^{\ell}(\iota)\right)(1 - \gamma_i)\kappa \\ 0 & \pi_2^r(\iota) + (1 - \pi_2^r(\iota))\kappa < \pi_2^{\ell}(\iota)(1 - \gamma_i) + \left(1 - \pi_2^{\ell}(\iota)\right)(1 - \gamma_i)\kappa \end{cases} \tag{3.5}$$

while the one of voters with negative bias ($\gamma_i < 0$) is given by:

$$\rho(\gamma_i, \iota) = \begin{cases} 1 & \pi_2^r(\iota)(1+\gamma_i) + (1-\pi_2^r(\iota))(1+\gamma_i)\kappa > \pi_2^\ell(\iota) + \left(1-\pi_2^\ell(\iota)\right)\kappa \\ x & \pi_2^r(\iota)(1+\gamma_i) + (1-\pi_2^r(\iota))(1+\gamma_i)\kappa = \pi_2^\ell(\iota) + \left(1-\pi_2^\ell(\iota)\right)\kappa \\ 0 & \pi_2^r(\iota)(1+\gamma_i) + (1-\pi_2^r(\iota))(1+\gamma_i)\kappa < \pi_2^\ell(\iota) + \left(1-\pi_2^\ell(\iota)\right)\kappa \end{cases} \tag{3.6}$$

where $x \in [0,1]$. An immediate consequence of (3.5) and (3.6) is that whenever a voter with positive bias (respectively, negative bias) votes for the candidate of party r (respectively, ℓ) with positive probability, all voters with higher (respectively, lower) bias will vote for the same candidate with certainty and viceversa. Further observe that the announcements of period 1 affect the electoral behavior only through their effect on $\left(\pi_2^\ell(.), \pi_2^r(.)\right)$.

We will show that the unique equilibrium of this game is uninformative: voters ignore the message sent by the candidates in period 1 and condition their electoral decision on signals only; consequently, candidates send uninformative messages and no information transmission takes place. We begin describing the behavior of voters:

Lemma 1 Let p_2^i represent the belief that voters assign to the candidate of party i being high valence at the beginning of period 2. Then there exists a threshold $\gamma\left(\kappa, p_2^\ell, p_2^r\right)$ such that all voters with bias higher (respectively, lower) than $\gamma\left(\kappa, p_2^\ell, p_2^r\right)$ vote for the candidate of the right (respectively, left) party. Furthermore $\gamma\left(\kappa, p_2^\ell, p_2^r\right) > 0$ if $p_2^\ell(\iota_S) > p_2^r(\iota_S)$ and $\gamma\left(\kappa, p_2^\ell, p_2^r\right) < 0$ if $p_2^\ell < p_2^r$.

Proof: Consider a candidate with positive bias, $\gamma_i > 0$. He will vote for the candidate of the right wing as long as:

$$p_2^r + \kappa(1-p_2^r) > p_2^\ell(1-\gamma_i) + \kappa\left(1-p_2^\ell\right)(1-\gamma_i)$$

or equivalently if

$$\gamma_i > \max\left\{0, \frac{\left(p_2^\ell - p_2^r\right)(1-\kappa)}{p_2^\ell(1-\kappa)+\kappa}\right\}$$

The first half of the proposition follows immediately observing that a symmetric result hold for a candidate with negative bias. To prove, the second part of the proposition, observe that $\frac{\left(p_2^\ell - p_2^r\right)(1-\kappa)}{p_2^\ell(1-\kappa)+\kappa} > 0$ $(-\frac{\left(p_2^r - p_2^\ell\right)(1-\kappa)}{p_2^r(1-\kappa)+\kappa} < 0)$ if and only if $p_2^\ell > (<)\, p_2^r$. ∎

Since in our model candidates want to maximize their plurality, they will try to move the threshold described in Lemma 1 in their favor. One of the instrument that candidates can use to achieve this goal is represented by electoral speeches, but, since voters are aware of candidates' incentives, these speeches will lack any credibility. The following Proposition provide a formal statement of this result.

Proposition 17 Suppose $\eta = 0$. Then, in every equilibrium, for every $\gamma_i \in [-1, 1]$ and $\forall m, m', m'', m''' \in M$, $\rho\left(\gamma_i, \left(\left(m, s^\ell\right), \left(m', s^r\right)\right)\right) = \rho\left(\gamma_i, \left(\left(m'', s^\ell\right), \left(m''', s^r\right)\right)\right)$.

Proof: Suppose, by the sake of contradiction, that the announcements in period 1 transmit some information that modify the electoral behavior of voters. Thus there is a party i and a pair of messages m_L and m_H such that $\pi^i(m_L) < q < \pi^i(m_H)$, where $\pi^i(m)$ denotes the probability voters assign to the candidate of party i being high valence after message $m \in M$. Let i^* be the party of the candidate who sends the message associated to the lowest probability of being high valence and let $\underline{m}$ be such a message. Then, there must exists a message $\overline{m}$ sent by the candidate of party i^* such that $\pi^{i^*}(\underline{m}) < q < \pi^{i^*}(\overline{m})$. Let us denote with $w(m_i, m_j)$ the effort exerted by the low valence candidate after that he sent message m_i and the other candidate sent message m_j. Observe that $\pi^{i^*}(\underline{m}) \neq 0$, since otherwise the candidate could improve his utility by sending message $\overline{m}$ and choosing distortion $w(\overline{m}, m^j) = \varepsilon$. Indeed, by choosing a sufficiently low ε, the low valence candidate would increase the expected mass of supporters by a positive amount whenever the signal sent is $s = 1$.[41] Therefore we conclude that $\pi^{i^*}(\underline{m}) \in (0, q)$ implying that the high candidate is

[41] To see this, observe that if voters listen to message $\overline{m}$ followed by signal $s = 1$, they will assign a probability higher than q to the candidate being high type. Let this probability be

sending message $\underline{m}$ with some probability. Furthermore the high valence candidate of party i^* is also sending some other message $\hat{m}$ with positive probability (otherwise $\pi(\underline{m}) \geq q$, contradicting our initial assumption). Optimality requires the expected support to be the same after these two messages and, since $\pi^{i^*}(\underline{m}) < q < \pi^{i^*}(\hat{m})$, this is possible only if there there exists some $\tilde{m}_j$ such that

$$\frac{\pi^{i^*}(\underline{m})}{\pi^{i^*}(\underline{m}) + (1 - \pi^{i^*}(\underline{m}))\, w_{i^*}(\underline{m}, \tilde{m}^j)} \geq \frac{\pi^{i^*}(\hat{m})}{\pi^{i^*}(\hat{m}) + (1 - \pi^{i^*}(\hat{m}))\, w_{i^*}(\hat{m}, \tilde{m}^j)}. \tag{3.7}$$

By definition of $\pi^{i^*}(\underline{m})$, we conclude that $\pi^{i^*}(\hat{m}) > \pi^{i^*}(\underline{m})$ and, consequently (3.7) implies $w(\hat{m}, \tilde{m}^j) > w(\underline{m}, \tilde{m}^j)$. Suppose $\pi(\hat{m}) = 1$, then (3.7) requires $w_{i^*}(\underline{m}, \tilde{m}^j) = 0$ contradicting optimality (the candidate would always prefer exerting an infinitesimal higher effort since he would gain some positive support by doing so). Thus $\pi(\hat{m}) \in \left(\pi^{i^*}(\underline{m}), 1\right)$ and consequently, message $\hat{m}$ is sent also by low valence candidates. Observe that (3.7) implies that the marginal benefit of exerting effort is greater after announcements $(\underline{m}, \tilde{m}^j)$ than after $(\hat{m}, \tilde{m}^j)$ and since the cost function is increasing and convex, we would need $w(\hat{m}, \tilde{m}^j) < w(\underline{m}, \tilde{m}^j)$ establishing the required contradiction.

We conclude that in every equilibrium, candidates' announcement do not convey any information concerning valence candidates. ∎

The next proposition characterizes the unique symmetric and uninformative equilibrium of the model. In this equilibrium candidates choose the same level of signal distortion after

π. If the candidate of the other party sends signal $s = 0$, the candidate of party i^* would get the support of all those voters biased in favor of party $j \neq i^*$, but whose bias has absolute value lower than $\left|\frac{(1-\kappa)\pi}{(1-\kappa)\pi+\kappa}\right|$. This will happen whenever the low type of party j exerts an effort lower than 1 after some pair of messages $(\overline{m}, m^j)$, $m^j \in M$. On the other hand, if the candidate exerts maximum effort after all pairs $(\overline{m}, m^j)$, there must exist at least one message after which the probability that the candidate of party i^* is high quality is lower or equal to q. In this case all voters biased in favor of party i^* and with a bias whose absolute value is lower than $\left|-\frac{(\pi-q)(1-\kappa)}{(1-\pi)\kappa+\kappa}\right|$ will support the candidate of party i^*.

any pair of messages.[42] We want to stress that although Proposition 18 characterizes the symmetric equilibrium only, Proposition 17 holds for every equilibrium.

Proposition 18 If $\eta = 0$, the unique symmetric equilibrium $\left(\left(t_i^*, w_i^*\right)_{i\in\{\ell,r\}}, \rho^*\right)$, $\pi\left(\left(t_i^*, w_i^*\right)_{i\in\{\ell,r\}}\right)$ is described as follows. For every $i \in \{\ell, r\}$, $t_i^* = t_i^{Un}$ and $w_i(q,\kappa,c) = \bar{w}$, where $\bar{w}$ is the unique fixed point of the following mapping

$$w = \frac{\left(F\left(\frac{(1-\kappa)q}{(1-\kappa w)q+\kappa w}\right) - \frac{1}{2}\right)}{c}.$$

Beliefs are determined by (3.1) and (3.2) and for every $i \in \{\ell, r\}$ and for every $m, m' \in M$

$$\rho^*(\gamma_i, ((m,0),(m',0))) = \rho^*(\gamma_i, ((m,1),(m',1))) = \begin{cases} 1 & \text{if } \gamma_i > 0 \\ x \in [0,1] & \text{if } \gamma_i = 0 \\ 0 & \text{if } \gamma_i < 0 \end{cases}$$

$$\rho^*(\gamma_i, ((m,0),(m',1))) = \begin{cases} 1 & \text{if } \gamma_i > -\frac{q(1-\kappa)}{q+\kappa(1-q)w(q,\kappa,c)} \\ x \in [0,1] & \text{if } \gamma_i = -\frac{q(1-\kappa)}{q+\kappa(1-q)w(q,\kappa,c)} \\ 0 & \text{if } \gamma_i < -\frac{q(1-\kappa)}{q+\kappa(1-q)w(q,\kappa,c)} \end{cases}$$

[42] Given the previous lemma, this is equivalent to saying that the candidates of the two parties choose the same level of signal distortion whenever voters have the same belief concerning their valence.

$$\rho^{*}\left(\gamma_i,\left(\left(m,1\right),\left(m',0\right)\right)\right)=\begin{cases}1 & if\ \gamma_i>\frac{q(1-\kappa)}{q+\kappa(1-q)w(q,\kappa,c)}\\ x\in[0,1] & if\ \gamma_i=\frac{q(1-\kappa)}{q+\kappa(1-q)w(q,\kappa,c)}\\ 0 & if\ \gamma_i<\frac{q(1-\kappa)}{q+\kappa(1-q)w(q,\kappa,c)}\end{cases}$$

Proof: Since Lemma 17 implies that candidates' announcements convey no information concerning their valence. Let these announcements be $\left(m^{\ell},m^{r}\right)$. We will prove the result proceeding backwards. After signal $s=0$, the voter assigns probability 0 to the agent having high valence. Therefore, if the other candidate also sent signal $s=0$, Lemma 1 implies that the threshold level is given by 0: all voters with positive (respectively, negative) bias will vote for the candidate of the right (respectively, left) party.

Suppose instead that the candidate of the right wing sent message $s=0$, while the one of the left party sent message $s=1$. Then, the threshold level will be given by:

$$\frac{\pi^{\ell}\left(1,m^{\ell},m^{r}\right)\left(1-\kappa\right)}{\pi^{\ell}\left(1,m^{\ell},m^{r}\right)\left(1-\kappa\right)+\kappa}>0$$

where

$$\pi^{\ell}\left(1,m^{\ell},m^{r}\right)=\frac{q}{q+\left(1-q\right)w^{\ell}\left(m^{\ell},m^{r}\right)}$$

Similarly, if the candidate of the right sent message $s=1$ and that of the left sent message $s=0$, Lemma 1 implies that the threshold level is given by

$$-\frac{\pi_{r}\left(1,m^{\ell},m^{r}\right)\left(1-\kappa\right)}{\pi_{r}\left(1,m^{\ell},m^{r}\right)\left(1-\kappa\right)+\kappa}$$

where

$$\pi_{r}\left(1,m^{\ell},m^{r}\right)=\frac{q}{q+\left(1-q\right)w^{r}\left(m^{\ell},m^{r}\right)}$$

Finally, if both agent send signal $s = 1$, the threshold level would be given by:

$$\gamma\left(m^{\ell}, m^{r}\right) = \begin{cases} \frac{\left(\pi^{\ell}\left(1,m^{\ell},m^{r}\right)-\pi^{r}\left(1,m^{\ell},m^{r}\right)\right)\cdot(1-\kappa)}{\pi^{\ell}\left(1,m^{\ell},m^{r}\right)\cdot(1-\kappa)+\kappa} & \text{if } \pi^{\ell}\left(1, m^{\ell}, m^{r}\right) > \pi^{r}\left(1, m^{\ell}, m^{r}\right) \\ 0 & \text{if } \pi^{\ell}\left(1, m^{\ell}, m^{r}\right) = \pi^{r}\left(1, m^{\ell}, m^{r}\right) \\ \frac{\left(\pi^{r}\left(1,m^{\ell},m^{r}\right)-\pi^{\ell}\left(1,m^{\ell},m^{r}\right)\right)\cdot(1-\kappa)}{\pi^{r}\left(1,m^{\ell},m^{r}\right)\cdot(1-\kappa)+\kappa} & \text{if } \pi^{\ell}\left(1, m^{\ell}, m^{r}\right) < \pi^{r}\left(1, m^{\ell}, m^{r}\right) \end{cases}$$

Consider the low valence candidate of the right party. The utility he gets from sending signal $s = 1$ is given by:

$$1 - \left(q + (1-q)\, w_{\ell}\left(m^{\ell}, m^{r}\right)\right) \cdot F\left(\gamma\left(m^{\ell}, m^{r}\right)\right) - \\ - (1-q)\left(1 - w_{\ell}\left(m^{\ell}, m^{r}\right)\right) \cdot F\left(-\frac{\pi^{r}\left(1, m^{\ell}, m^{r}\right)(1-\kappa)}{\pi^{r}\left(1, m^{\ell}, m^{r}\right)(1-\kappa)+\kappa}\right)$$

while his utility from having signal $s = 0$ sent is equal to:

$$1 - (1-q)\left(1 - w_{\ell}\left(m^{\ell}, m^{r}\right)\right) \cdot F(0) - \\ - \left(q + (1-q)\, w_{\ell}\left(m^{\ell}, m^{r}\right)\right) \cdot F\left(\frac{\pi^{\ell}\left(1, m^{\ell}, m^{r}\right)(1-\kappa)}{\pi^{\ell}\left(1, m^{\ell}, m^{r}\right)(1-\kappa)+\kappa}\right)$$

Thus the marginal benefit of additional signal distortion for the candidate of party r at information set $\left(m^{\ell}, m^{r}\right)$ is given by:

$$\left(q + (1-q)\, w_{\ell}\left(m^{\ell}, m^{r}\right)\right) \cdot \left(F\left(\frac{\pi^{\ell}\left(1, m^{\ell}, m^{r}\right)(1-\kappa)}{\pi^{\ell}\left(1, m^{\ell}, m^{r}\right)(1-\kappa)+\kappa}\right) - F\left(\gamma\left(m^{\ell}, m^{r}\right)\right)\right) + \\ + (1-q)\left(1 - w_{\ell}\left(m^{\ell}, m^{r}\right)\right) \cdot \left(\frac{1}{2} - F\left(-\frac{\pi^{r}\left(1, m^{\ell}, m^{r}\right)(1-\kappa)}{\pi^{r}\left(1, m^{\ell}, m^{r}\right)(1-\kappa)+\kappa}\right)\right)$$

Similarly, the benefit from an additional unit of signal distortion for the candidate of party ℓ is given by:

$$\left(q+(1-q)\,w_r\left(m^\ell,m^r\right)\right)\cdot\left(F\left(\gamma\left(m^\ell,m^r\right)\right)-F\left(-\frac{\pi^r\left(1,m^\ell,m^r\right)(1-\kappa)}{\pi^r\left(1,m^\ell,m^r\right)(1-\kappa)+\kappa}\right)\right)+$$
$$+(1-q)\left(1-w_r\left(m^\ell,m^r\right)\right)\cdot\left(F\left(\frac{\pi^\ell\left(1,m^\ell,m^r\right)(1-\kappa)}{\pi^\ell\left(1,m^\ell,m^r\right)(1-\kappa)+\kappa}\right)-\frac{1}{2}\right)$$

The marginal cost of signal distortion for agent i is given by cw_i. Thus the equilibrium effort level is the solution to these system of two equations:

$$cw_r=(1-q)\left(1-w_\ell\left(m^\ell,m^r\right)\right)\cdot\left(\frac{1}{2}-F\left(-\frac{\pi^r\left(1,m^\ell,m^r\right)(1-\kappa)}{\pi^r\left(1,m^\ell,m^r\right)(1-\kappa)+\kappa}\right)\right)+$$
$$+\left(q+(1-q)\,w_\ell\left(m^\ell,m^r\right)\right)\cdot\left(F\left(\frac{\pi^\ell\left(1,m^\ell,m^r\right)(1-\kappa)}{\pi^\ell\left(1,m^\ell,m^r\right)(1-\kappa)+\kappa}\right)-F\left(\gamma\left(m^\ell,m^r\right)\right)\right)$$

and

$$cw_\ell=(1-q)\left(1-w_r\left(m^\ell,m^r\right)\right)\cdot\left(F\left(\frac{\pi^\ell\left(1,m^\ell,m^r\right)(1-\kappa)}{\pi^\ell\left(1,m^\ell,m^r\right)(1-\kappa)+\kappa}\right)-\frac{1}{2}\right)+$$
$$+\left(q+(1-q)\,w_r\left(m^\ell,m^r\right)\right)\cdot\left(F\left(\gamma\left(m^\ell,m^r\right)\right)-F\left(-\frac{\pi^r\left(1,m^\ell,m^r\right)(1-\kappa)}{\pi^r\left(1,m^\ell,m^r\right)(1-\kappa)+\kappa}\right)\right)$$

Imposing symmetry, we get $w_\ell\left(m^r,m^\ell\right)=w_r\left(m^r,m^\ell\right)=w\;\forall\left(m^\ell,m^r\right)$ so that $\gamma\left(m^\ell,m^r\right)=0$ and the equilibrium signal distortion is defined by the following equation:

$$w=\frac{F\left(\frac{\pi_1(1-\kappa)}{\pi_1(1-\kappa)+\kappa}\right)-\frac{1}{2}}{c}$$

or equivalently:

$$w=\frac{F\left(\frac{q(1-\kappa)}{q(1-w\kappa)+w\kappa}\right)-\frac{1}{2}}{c}$$

It is easy to check that Assumption 8 together with Brower fixed point theorem immediately implies the existence of a fixed point in $(0,1)$. Furthermore, the derivative of the left hand side is negative on the whole domain and therefore the fixed point is unique. Let us denote this fixed point with $w(\Pi, q, \kappa, c)$. Therefore $\pi(1) = \frac{q}{q+(1-q)w(\Pi,q,\kappa,c)}$. The remaining of the proposition follows immediately from Lemmata 1 and 17. ∎

Thus, if voters are standard expected utility maximizer, the unique symmetric equilibrium is uninformative and it involves a positive amount of signal distortion. Intuitively, in a model with standard utilities, candidates' announcements affect the behavior of voters only through their effect on beliefs. Thus, low valence candidates will try to distort beliefs in order to gain the support of the highest mass of voters possible; this will result in a lack of credibility and will induce politicians to exert a positive effort in signal distortion.

3.4.2 Voters with Reference-Dependent Utility

Now, we will assume that voters have reference-dependent preferences and exhibit loss aversion (that is, Assumption 9 holds) and we will show that a fully informative equilibrium coexists with an uninformative one.

The Fully Informative Equilibrium

In the model without reference dependence politicians' speeches affect voters' behavior only through their effect on belief at the electoral stage. As shown in Section 3.4.1, this channel is insufficient to lead to any type of information transmission. Indeed, in our model, voters lack any mechanism to commit themselves to punish a candidate who lied: were a lie be discovered, a voter's ranking of candidates would not change. Therefore lies would represent a riskless opportunity to modify electoral behavior, and candidates will always be willing to make false announcements in the hope that their lie goes undetected and that the support in the population increases.

Reference dependence and loss aversion overcome this commitment problem by introducing an additional link between period-0 announcements and period-2 electoral behavior: the effect on reference utility. In the setting of our model, this additional channel will induce a positive mass of biased voters to "punish" a liar by voting for the opposing candidate. In this section, we will show that this type of behavior can lead to full information transmission for a range of parameters that vary with the degree of loss aversion. We proceed in two steps: we first analyze the behavior of voters assuming that they believe that candidates announced their valence truthfully; then we focus our attention on candidates taking as given the behavior of voters.

Voters' Behavior We start observing that in a fully informative equilibrium the message space used by candidates can be assumed to be equal to $M = \{\theta_L, \theta_H\}$ without any loss of generality.[43] Thus, the information sets are represented by:

$$\begin{aligned} \mathcal{I}_A^{FI} &= \left\{ \left(m^\ell, m^r\right) : m_\ell, m_r \in \{\theta_L, \theta_H\} \right\} \\ \mathcal{I}_S^{FI} &= \left\{ \left(\left(m^\ell, s^\ell\right), \left(m^r, s^r\right) \right) : m^\ell, m^r \in \{\theta_L, \theta_H\}, s^\ell, s^r \in \{0,1\} \right\} \end{aligned}$$

Let $t_i^{Tr}(.)$ be the truthful strategy of agent i. Then 3.1 immediately implies that for every $i, j \in \{\ell, r\}$ with $i \neq j$ and for every $m^j \in \{\theta_L, \theta_H\}$:

$$\begin{aligned} \pi_1^i \left(\left(\theta_H, m^j\right) \mid \left(t_i^{Tr}, w_i\right)_{i \in \{\ell,r\}} \right) &= 1, \\ \pi_1^i \left(\left(\theta_L, m^j\right) \mid \left(t_i^{Tr}, w_i\right)_{i \in \{\ell,r\}} \right) &= 0. \end{aligned}$$

In words, in a truthful equilibrium a candidate will assign probability 1 to the valence of the candidate being the one announced. Therefore the voter will expect to receive signal 1 from

[43] In particular, we can assume that any message different from "θ_H" is interpreted by the agents as message "θ_L".

any candidate who announced to be high valence. Formally, for every $(m^j, s^j) \in M \times \{0,1\}$

$$\chi\left(\left(\left(\theta_H, 0\right), \left(m^j, s^j\right)\right) \mid \left(\theta^H, m^j\right), \left(t_i^{Tr}, w_i\right)_{i\in\{\ell,r\}}\right) = 0.$$

The characterization of voters' equilibrium behavior will require a two steps-procedure. We will first characterize the set of voters' dynamic consistent strategies, namely those strategies that would be optimal if the reference utility of voters were determined based on them. Then, in case of multiple dynamic consistent strategies, we will select the one that maximize the total utility of the voters. This analysis is summarized in the next proposition:

Proposition 19 Suppose Assumption 9 holds. Let $\rho^*(.,.)$ be the optimal dynamic consistent strategy given $\left(t_i^{Tr}, w_i\right)_{i\in\{\ell,r\}}$. Then for each information set $\iota_S \in I_S^{FI}$, there exists a threshold level $\hat{\gamma}(\iota_S)$ such that:

$$\rho^*(\gamma_i, \iota_S) = \begin{cases} 1 & \text{if } \gamma_i > \hat{\gamma}(\iota_S) \\ x \in [0,1] & \text{if } \gamma_i = \hat{\gamma}(\iota_S) \\ 0 & \text{if } \gamma_i < \hat{\gamma}(\iota_S) \end{cases}$$

Furthermore the threshold level $\hat{\gamma}(\iota_S)$ is characterized as follows:

$$\hat{\gamma}(\iota_S) = \begin{cases} -\gamma^*(\kappa) & \text{if } \iota_S = ((\theta_H, 0), (\theta_H, 1)) \\ -\gamma^{**}(\kappa, \eta, \lambda) & \text{if } \iota_S = ((\theta_H, 1), (\theta_L, 1)) \text{ or } \iota_S = ((\theta_L, 0), (\theta_H, 1)) \\ \gamma^*(\kappa) & \text{if } \iota_S = ((\theta_H, 1), (\theta_H, 0)) \\ \gamma^{**}(\kappa, \eta, \lambda) & \text{if } \iota_S = ((\theta_H, 1), (\theta_L, 1)) \text{ or } \iota_S = ((\theta_H, 1), (\theta_L, 0)) \\ 0 & \text{otherwise} \end{cases}$$

where

$$\gamma^*(\kappa) = 1 - \kappa, \ \gamma^{**}(\kappa, \eta, \lambda) = \frac{(1-\kappa)(1+\eta)}{1 + (1-\kappa)\eta + \kappa\eta\lambda}$$

Proof: See Section 3.6.1.

Although the formal proof of proposition 19 is in the appendix, we provide its main intuition here. The first step in the proof is the characterization of the reference utility associated with dynamic consistent strategies for each possible pair of announcements $\left(m^\ell, m^r\right)$. We start characterizing reference utilities for two reasons: on the one hand, the focus on reference utilities instead of reference points simplifies notation; on the other hand, this choice allows us to specify the optimal behavior of voters only at those information sets that play a role in determining voters' reference utility, namely those information sets that voters believe to occur with positive probability.[44] In the proof, we show that the reference utility will always be uniquely determined for voters with strong or weak ideological biases, but will take multiple values if voters are moderately biased ($\gamma_i \in [\gamma^{**}(\kappa, \eta, \lambda), \gamma^*(\kappa)]$ or

[44]The actual characterization of the optimal dynamic strategy is not affected by the fact that voters assign zero probability to some node and all the results would still hold if we assing a positive, but small probability to these nodes.

$\gamma_i \in [-\gamma^*(\kappa), -\gamma^{**}(\kappa,\eta,\lambda)]$) and candidates announced (θ_H, θ_L) or (θ_L, θ_H). Intuitively, people with extreme biases will always vote based on ideology, while people with no bias will always vote based on valence only (and in a truthful equilibrium they will assign probability 1 to the valence announced by candidates). However, people with moderate degrees of ideological bias, will have a unique dynamic consistent strategy only if the ranking between candidate is clear (this happens if the candidate of their preferred party is high valence or if both candidates are low valence); in the remaining case, there will be a dynamic consistent strategy that prescribes to vote for the left-wing candidate and a different one that prescribes to vote for the right-wing candidate (in addition there will also be a dynamic consistent strategy that prescribes a mixed electoral behavior). This multiplicity in dynamic consistent strategies translates into a multiplicity of reference utilities; the second part of the proof selects the utility associated with the dynamic consistent strategy that maximizes voters' total utility.[45] Finally, in the third and last part of the proof, we exploit the characterization of reference utility to provide a full description of the optimal dynamic consistent strategy.

Proposition 19 has important implications on the behavior of voters. Consider the low valence candidate of party r and assume that the candidates of the other party are playing a truthful communication strategy and that lies are detected with some exogenous probability.[46] Proposition 19 implies that if he were to announce his valence truthfully, he would get the support of half of the electorate with probability $1-q$ and of a mass $1-F(\gamma^{**}(\kappa,\eta,\lambda))$ of the voters with probability q.[47] These masses of voters are represented respectively by the blue and red solid lines in the bottom part of Figure 2. Suppose, instead that he lies pretending to be a high valence candidate. If his lie goes undetected (that is, the signal does not falsify him), he will get the support of of a mass of voters equal to $1-F(-\gamma^{**}(\kappa,\eta,\lambda))$

[45] In this case, the set of biases for which multiple dynamic consistent strategies are possible, will have probability 0.

[46] We will address the issue of signal distortion in the next section.

[47] Recall that q is the probability with which the candidate of the other party has high valence

with probability $1-q$ and of half of the electorate with probability q. However if his lies are discovered, the support would shrink to $\frac{1}{2}$ and $1-F(\gamma^*(\kappa))$ resepctively. These masses of voters are represented in the top part of Figure 2, where the dashed line represents those voters who would vote for the candidate of party r only if they do not detect a lie. Since $\gamma(\kappa^*) > \gamma^{**}(\kappa,\eta,\lambda)$, we can see that a false announcement may decrease the support in favour of the right-wing candidate to a level lower than the one the candidate could have obtained by telling the truth (of course, if the lie is not detected, the opposite would be true). This mechanism is what induces candidates to announce their valence truthfully. The characterization of Proposition 19 depends on two thresholds levels: $\gamma^*(\kappa)$, $\gamma^{**}(\kappa,\eta,\lambda)$. In the following remark we summarize the dependency of these thresholds on the other parameters of the model.

Remark 4 (i) $\gamma^*(\kappa)$ is decreasing in κ.
(ii) $\gamma^{**}(\kappa,\eta,\lambda) = \frac{(1-\kappa)(1+\eta)}{1+(1-\kappa)\eta+\kappa\eta\lambda}$ is decreasing in κ, η and λ. Furthermore $\gamma^{**}(\kappa,\eta,1) = 1-\kappa$ and $\lim_{\lambda\to\infty}\gamma^{**}(\kappa,\eta,\lambda) = 0$.

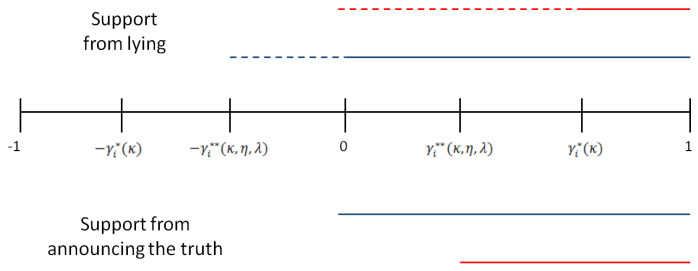


Figure 8: Support for type (r,θ_L).

Red/Blue Line: expected support if the opponent is high/low valence.

Dashed Lines: voters who support the candidate only if the lie is not detected.

Thus, by inspection of Figure 8, we can conclude that if $\lambda = 1$, candidates' expected loss from lying is zero, while the expected gain is positive. On the opposite, if $\lambda \to \infty$, the expected loss from lying increases and the expected gain decreases. This suggests that the incentives to tell the truth increase with the degree of loss aversion. To make this intuition precise, and to relax the assumption that lies are detected with some exogenous probability, we have to analyze the behavior of candidates.

Candidates' Behavior Recall that in a truthful equilibrium, the behavior of the candidates of party i can be described by a function $t_i^{Tr} : \{\theta_L, \theta_H\} \to [0,1]$ representing the communication strategy of party i and by a function $w_i : \mathcal{I}_A \to [0,1]$ representing the degree of signal distortion chosen by the low valence candidate of party i. We start analyzing the effort decision of the low valence candidate at different information sets. Suppose that this candidate has announced his type truthfully in period 1; then voters will keep believing that the candidate is low valence regardless of the actual signal received. Thus, any effort to distort the signal will be pure waste. On the other hand, if type (i, θ_L) lied about his valence, the previous discussion implies that he will have an incentive to distort the signal to avoid losing the dashed mass of voters in Figure 8. The following Proposition characterizes the optimal signal distortion.

Proposition 20 Suppose Assumption 9 holds and that voters follow the optimal dynamic consistent strategy characterized in Proposition 19. Then in a truthful equilibrium, the optimal effort level chosen by low valence candidates is given by:

$$w_i^*\left(m^i, m^j\right) = \begin{cases} \frac{F(\gamma^*(\kappa,\eta,\lambda))}{c} - \frac{1}{2c} & \text{if } (m^i, m^j) = (\theta_H, \theta_H) \\ \frac{F(\gamma^{**}(\kappa,\eta,\lambda))}{c} - \frac{1}{2c} & \text{if } (m^i, m^j) = (\theta_H, \theta_L) \quad \forall i, j \in \{\ell, r\},\ i \neq j. \\ 0 & \text{otherwise} \end{cases}$$

Proof: Consider the candidate of party r. In a truthful equilibrium in which the candidate of the right party announced $m_r = \theta_L$, the beliefs of voters will be given by $\pi_2^r(\iota_S) = 0$ regardless of the actual signal received. Thus, signal distortion has a positive cost and cannot affect the electoral behavior. We conclude that:

$$w_r(\theta_L, \theta_L) = w_r(\theta_H, \theta_L) = 0.$$

Now, suppose that the low candidate lied about his valence pretending to be θ_H. Thus, Propositions 2 and 3 imply that all voters with positive bias will have reference point equal to δ_1. Furthermore, the same Propositions imply that if the candidate of party ℓ declared to be low valence, all voters with bias in $\gamma_i \in [-\gamma^{**}(\kappa, \eta, \lambda), 0]$ would have reference point given by $(1+\gamma_i)[1] + \gamma_i[0]$; on the opposite, if the candidate of party ℓ announced to be a high valence, all voters with negative bias would have reference point given by δ_1. Moreover, the actual behavior of voters for each possible information set is described by Proposition 19.

In a truthful equilibrium, the candidates of party r believes that the candidate of party ℓ reported his valence truthfully, so they will assign probability 1 (respectively, 0) to the candidate having high valence after announcement $m^\ell = \theta_H$, (respectively, $m^\ell = \theta_L$).

Therefore, after announcements (θ_L, θ_H), the benefit from sending message $s = 1$ instead of $s = 0$ is given by $(1 - F(-\gamma^{**}(\kappa, \eta, \lambda)) - 1 + F(0))$. Using the fact that the distribution of $F(.)$ is symmetric, this benefit can be rewritten as:

$$B((\theta_L, \theta_H)) = F(\gamma^{**}(\kappa, \eta, \lambda)) - \frac{1}{2}$$

Therefore at information set (θ_L, θ_H), the candidate of party r would choose w to solve:

$$\max_{w \in [0,1]} w\left(F(\gamma^{**}(\kappa, \eta, \lambda)) - \frac{1}{2}\right) - \frac{cw^2}{2}$$

Then Assumption 8 implies that:

$$w_r^*\left(\left(\theta_L,\theta_H\right)\right)=\frac{F\left(\gamma^{**}\left(\kappa,\eta,\lambda\right)\right)}{c}-\frac{1}{2c}$$

Similarly, after announcements $\left(\theta_H,\theta_H\right)$, the benefit from sending signal $s=1$ is given by:

$$B\left(\left(\theta_H,\theta_H\right)\right)=F\left(\gamma^*\left(\kappa\right)\right)-\frac{1}{2}$$

Consequently, it is easy to show that:

$$w_r^*\left(\left(\theta_H,\theta_H\right)\right)=\frac{F\left(\gamma^*\left(\kappa\right)\right)}{c}-\frac{1}{2c}$$

The analysis of the effort exerted by the low valence of party ℓ is identical and omitted. ∎

Intuitively, the previous proposition states the obvious result that the optimal level of signal distortion is the one that equates the marginal benefit from increasing the probability of sending signal $s=1$ with the marginal cost associated with signal distortion. Observe that the marginal benefit of signal distortion for the candidate of party i changes with the announcement made by the candidate of party j. This reflects the fact that in a truthful equilibrium the belief of voters concerning the valence of a candidate is affected by the announcement he made in period 0.

We are now ready to analyze the communication strategy of the candidates and to verify that they will be willing to announce their valence truthfully. The next proposition provides conditions under which this will indeed be the case; these conditions are stated in terms of cost associated with signal distortion. However, we can interpret them as relating the probability of being caught with the degree of loss aversion: as long as the cost associated with signal distortion is sufficiently high, candidates will be unable to prevent voters from finding out their true valence with some probability and to react to lies in the way described before.

Proposition 21 Suppose Assumption 9 holds. There exists a $c^*(q,\kappa,\eta,\lambda) \geq \frac{1}{2}$ such that as long as $c > c^*(q,\kappa,\eta,\lambda)$, a fully informative equilibrium exists. In this equilibrium for every $i \in \{\ell, r\}$, $t_i = t_i^{Tr}$, w_i and ρ^* are characterized by Propositions 20 and 19 and beliefs are determined by 3.1 and 3.2. Furthermore $c^*(q,\kappa,\eta,\lambda)$ is decreasing in λ and there exists $\lambda^*(q,\kappa,\eta) > 1$ such that if $\lambda > \lambda^*(q,\kappa,\eta)$, $c^*(q,\kappa,\eta,\lambda) = \frac{1}{2}$.

Proof: Consider the candidate of the right party and suppose he believes that the candidate of party ℓ reveals his valence truthfully (the analysis for party ℓ's candidate is equivalent and omitted).

Consider type θ_H first. If he announces his valence truthfully, the previous propositions imply that the expected support of the candidate would be:

$$q \cdot (1 - F(0)) + (1-q) \cdot F(\gamma^{**}(\kappa,\eta,\lambda))$$

On the other hand if he lies and claims to be a low skilled candidate, the expected mass of voters who will support him will be given by:

$$q \cdot (1 - F(\gamma^{**}(\kappa,\eta,\lambda))) + (1-q) \cdot (1 - F(0))$$

Observe that:

$$\begin{gathered} q \cdot (1 - F(0)) + (1-q) \cdot F(\gamma^{**}(\kappa,\eta,\lambda)) \geq \\ \geq q \cdot (1 - F(\gamma^{**}(\kappa,\eta,\lambda))) + (1-q) \cdot (1 - F(0)) \\ \Longleftrightarrow \\ F(\gamma^{**}(\kappa,\eta,\lambda)) \geq \frac{1}{2} \end{gathered}$$

Then $\gamma^{**}(\kappa,\eta,\lambda) > 0$ and $F(.)$ symmetric around 0 imply $t_r(\theta_H)[\theta_H] = 1$.

Now consider a candidate with valence θ_L. If he announces his valence truthfully, Proposition

19 implies that his expected support would be given by:

$$q \cdot (1 - F(\gamma^{**}(\kappa, \eta, \lambda))) + (1 - q) \cdot (1 - F(0))$$

On the other hand, if he lies and announces that his valence is θ_H, his utility will be equal to:

$$\begin{aligned} qw(\theta_H, \theta_H)(1 - F(0)) + q(1 - w(\theta_H, \theta_H))(1 - F(\gamma^*(\kappa))) - \\ - q\frac{c(w(\theta_H, \theta_H))^2}{2} + (1 - q)w(\theta_L, \theta_H)(1 - F(-\gamma^{**}(\kappa, \eta, \lambda))) + \\ + (1 - q)(1 - w(\theta_L, \theta_H))(1 - F(0)) - (1 - q)\frac{c(w(\theta_L, \theta_H))^2}{2} \end{aligned}$$

Substituting the optimal value of effort and rearranging terms, we get that the low candidate will tell the truth as long as:

$$\begin{aligned} q(F(\gamma^*(\kappa)) - F(\gamma^{**}(\kappa, \eta, \lambda))) > q\left(\frac{(F(\gamma^*(\kappa)))^2}{2c} - \frac{F(\gamma^*(\kappa))}{2c} + \frac{1}{8c}\right) + \\ + (1 - q)\left(\frac{(F(\gamma^{**}(\kappa, \eta, \lambda)))^2}{2c} - \frac{F(\gamma^{**}(\kappa, \eta, \lambda))}{2c} + \frac{1}{8c}\right) \end{aligned} \tag{3.8}$$

Note that as $c \to \infty$, the right hand side of the inequality converges to 0 and since $F(\gamma^*(\kappa)) > F(\gamma^{**}(\kappa, \eta, \lambda))$, the inequality is satisfied.

$$\begin{aligned} q(F(\gamma^*(\kappa)) - F(\gamma^{**}(\kappa, \eta, \lambda))) > q\left((F(\gamma^*(\kappa)))^2 - F(\gamma^*(\kappa)) + \frac{1}{4}\right) + \\ + (1 - q)\left((F(\gamma^{**}(\kappa, \eta, \lambda)))^2 - F(\gamma^{**}(\kappa, \eta, \lambda)) + \frac{1}{4}\right) \end{aligned}$$

On the other hand if $c = \frac{1}{2}$, the right hand side is equal to:

$$q\left(\left(F\left(\gamma^{*}\left(\kappa\right)\right)\right)^{2} - F\left(\gamma^{*}\left(\kappa\right)\right) + \frac{1}{4}\right) + \\ + \left(1-q\right)\left(\left(F\left(\gamma^{**}\left(\kappa,\eta,\lambda\right)\right)\right)^{2} - F\left(\gamma^{**}\left(\kappa,\eta,\lambda\right)\right) + \frac{1}{4}\right)$$

and the previous inequality is satisfied if

$$q\left(\left(F\left(\gamma^{*}\left(\kappa\right)\right)\right)^{2} - 2F\left(\gamma^{*}\left(\kappa\right)\right) - \left(F\left(\gamma^{**}\left(\kappa,\eta,\lambda\right)\right)\right)^{2} + 2F\left(\gamma^{**}\left(\kappa,\eta,\lambda\right)\right)\right) \\ + \left(F\left(\gamma^{**}\left(\kappa,\eta,\lambda\right)\right)\right)^{2} - F\left(\gamma^{**}\left(\kappa,\eta,\lambda\right)\right) + \frac{1}{4} < 0$$

or equivalently:

$$q > \frac{\left(F\left(\gamma^{**}\left(\kappa,\eta,\lambda\right)\right)\right)^{2} - F\left(\gamma^{**}\left(\kappa,\eta,\lambda\right)\right) + \frac{1}{4}}{\left(2F\left(\gamma^{*}\left(\kappa\right)\right) - \left(F\left(\gamma^{*}\left(\kappa\right)\right)\right)^{2} - 2F\left(\gamma^{**}\left(\kappa,\eta,\lambda\right)\right) + \left(F\left(\gamma^{**}\left(\kappa,\eta,\lambda\right)\right)\right)^{2}\right)}$$

Define the right hand side of the inequality $q^{*}\left(\kappa,\eta,\lambda\right)$. Observe that both the numerator and the denominator are always positive. Furthermore, we know that if $\lambda \to \infty$, $\gamma^{**}\left(\kappa,\eta,\lambda\right) \to 0$ and consequently $F\left(\left(\gamma^{**}\left(\kappa,\eta,\lambda\right)\right)\right) \to \frac{1}{2}$. In this case $q^{*}\left(\kappa,\eta,\lambda\right) \to 0$ guaranteeing that the inequality would be satisfied for any positive q.

Furthermore the right hand side of 3.8 is decreasing in c and converges to ∞ and to 0 as $c \to 0$ and $c \to \infty$, respectively. The left hand side, instead, is independent on c and it is positive. Therefore there exists a unique $c^{*}\left(q,\kappa,\eta,\lambda\right)$ such that the two sides of 3.8 are equal and if $c > c^{*}\left(q,\kappa,\eta,\lambda\right)$, the candidate will prefer to reveal his valence truthfully. The value $c^{*}\left(q,\kappa,\eta,\lambda\right)$ is implicitly defined by the following equation:

$$q\left(F^{*}\left(\gamma^{*}\left(\kappa\right)\right) - F\left(\gamma^{**}\left(\kappa,\eta,\lambda\right)\right)\right) = q\left(\frac{\left(F\left(\gamma^{*}\left(\kappa\right)\right)\right)^{2}}{2c} - \frac{F\left(\gamma^{*}\left(\kappa\right)\right)}{2c} + \frac{1}{8c}\right) + \\ + \left(1-q\right)\left(\frac{\left(F\left(\gamma^{**}\left(\kappa,\eta,\lambda\right)\right)\right)^{2}}{2c} - \frac{F\left(\gamma^{**}\left(\kappa,\eta,\lambda\right)\right)}{2c} + \frac{1}{8c}\right)$$

Applying the implicit function theorem, we conclude that the derivative of $c^*(q,\kappa,\eta,\lambda)$ with respect to λ :

$$-\frac{-q\cdot\frac{\partial F(\gamma^{**})}{\partial\gamma}\cdot\frac{\partial\gamma^{**}}{\partial\lambda}+\frac{(1-q)}{2c}\cdot\frac{\partial F(\gamma^{**})}{\partial\gamma}\cdot\frac{\partial\gamma^{**}}{\partial\lambda}-\frac{1-q}{c}\cdot F(\gamma^{**})\cdot\frac{\partial F(\gamma^{**})}{\partial\gamma}\cdot\frac{\partial\gamma^{**}}{\partial\lambda}}{q\left(\frac{(F(\gamma^{*}(\kappa)))^2}{2c^2}-\frac{F(\gamma^{*}(\kappa))}{2c^2}+\frac{1}{8c^2}\right)+(1-q)\cdot\left(\frac{(F(\gamma^{**}(\kappa,\eta,\lambda)))^2}{2c^2}-\frac{F(\gamma^{**}(\kappa,\eta,\lambda))}{2c^2}+\frac{1}{8c^2}\right)}$$

Observe that the denominator of this expression is positive, so that the sign of this derivative will be equal to the sign of:

$$q\cdot\frac{\partial F(\gamma^{**})}{\partial\gamma}\cdot\frac{\partial\gamma^{**}}{\partial\lambda}-\frac{(1-q)}{2c}\cdot\frac{\partial F(\gamma^{**})}{\partial\gamma}\cdot\frac{\partial\gamma^{**}}{\partial\lambda}+\frac{1-q}{c}\cdot F(\gamma^{**})\cdot\frac{\partial F(\gamma^{**})}{\partial\gamma}\cdot\frac{\partial\gamma^{**}}{\partial\lambda}\qquad(3.9)$$

Since $\frac{\partial F(\gamma^{**})}{\partial\gamma}>0$ and $\frac{\partial\gamma^{**}}{\partial\lambda}<0$, 3.9 is negative as long as

$$F(\gamma^{**}(\kappa,\eta,\lambda))>\frac{1}{2}-c\cdot\frac{q}{1-q}$$

which is always the case since $\gamma^{**}(\kappa,\eta,\lambda)>0$ and $F(0)=\frac{1}{2}$. Now, the statement of the Proposition follows from the previous characterization. ∎

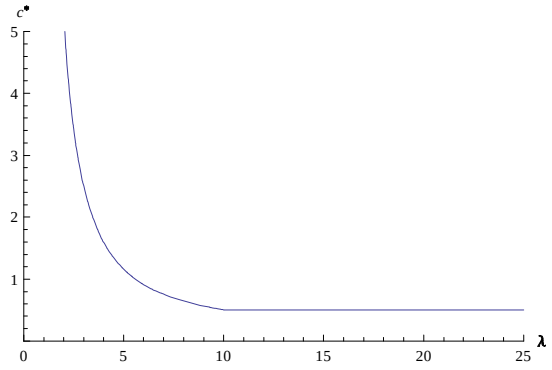


Figure 9: $c^*(q,\kappa,\eta,\lambda)$ as a function of λ. when $q=0.2$, $\kappa=0.4$ and $\eta=0.5$ when $F(.)$ is uniform on $[-1,1]$.

Proposition 21 states that if the cost associated with signal distortion is not too low, a fully revealing equilibrium will indeed exists. Furthermore, this threshold value for the cost of signal distortion is decreasing in the degree of loss aversion and reaches $\frac{1}{2}$ if the coefficient of loss aversion is sufficiently high.[48] Indeed, by the discussion of Section 3.4.2, an increase in loss aversion has the double effect of increasing the expected cost from lying and decreasing its expected benefit,[49] and will thus decrease the required probability for lie detection. The actual shape of the relationship between $c^*(q, \kappa, \eta, \lambda)$ and λ will depend on the specific empirical frequency of voters, $F(.)$. Figure 9, plot $c^*(q, \kappa, \eta, \lambda)$ as a function of λ in the special case in which ideological biases are uniformly distributed in $[-1, 1]$; the shape of $c^*(q, \kappa, \eta, \lambda)$ as a function of q, κ and η is similar.

The Uninformative Equilibrium

Although reference-dependent utility may lead to information transmission, it does not rule out the possibility of an uninformative equilibrium. This reflects a standard feature of the literature on strategic communication: uninformative (babbling) equilibria can always be supported by the assumptions that voters do not trust candidates' statements and ignore them. If this were to happen, candidates will have no incentive to announce their type sincerely and the uninformative strategy will be trivially optimal.

In this section, we will show that uninformative equilibria do exist. In an uninformative equilibrium, the actual message sent by the candidates does not play any role in determining voters' behavior; thus we can focus our attention on a single pair of announcements (m, m) at which $\pi_1^i\left((m, m) \mid \left(t_i^{Un}, w_i\right)_{i\in\{\ell,r\}}\right) = q$ for every $i \in \{\ell, r\}$. Given that the actual announcement does not play any role in the analysis, we will omit to specify the dependency of beliefs and strategies on the communication strategy $\left(t_i^{Un}\right)_{i\in\{\ell,r\}}$, information set (m, m)

[48] Recall that Assumption 8 requires the cost to be higher than $\frac{1}{2}$.

[49] Observe that to make this argument complete, we have to take into account the effect of a change in λ on the signal distortion effort.

and belief $\pi_1^i\left(m, m \mid \left(t_i^{Un}, w_i\right)_{i\in\{\ell,r\}}\right)$. Note that in an uninformative equilibrium:

$$\pi_2^i\left(s^\ell, s^r \mid (w_i)_{i\in\{\ell,r\}}\right) = \begin{cases} 0 & \text{if } s^i = 0 \\ & , \ i \in \{\ell, r\} \\ \frac{q}{q+(1-q)w_i} & \text{if } s^i = 1 \end{cases}$$

and consequently:

$$\chi\left(s^\ell, s^r\right) = \begin{cases} (1-q)(1-w_\ell)(1-q)(1-w_r) & \left(s^\ell, s^r\right) = (0,0) \\ (1-q)(1-w_\ell)((1-q)w_r+q) & \left(s^\ell, s^r\right) = (0,1) \\ ((1-q)w_\ell+q)(1-q)(1-w_r) & \left(s^\ell, s^r\right) = (1,0) \\ ((1-q)w_\ell+q)((1-q)w_r+q) & \left(s^\ell, s^r\right) = (1,1) \end{cases} . \tag{3.10}$$

In the following Proposition, we prove the existence of a symmetric and uninformative equilibrium in which voters choose the same level of signal distortion, so that $w_\ell = w_r = w$.

Proposition 22 Suppose Assumption 9 holds. In the model with reference dependent utility, there exists a symmetric and uninformative equilibrium.

Proof: See Section 3.6.2.

3.5 Conclusion

In this paper we provide a model to explain two apparently contradictory claims: (i) politicians' electoral speeches are not credible, but (ii) politicians are held accountable for their electoral promises. To this goal, we build a model of electoral competition in which two parties compete to get the support of a mass of voters who care about candidates' valence

and party affiliation. In our model, if voters care about material utility only, politicians' announcements are uninformative: since politicians always have an incentive to pretend to be high valence, their statements will lack any credibility and voters will ignore them. The introduction of reference dependence and loss aversion overcome this problem by adding an additional channel through which politicians' announcements affect voters behavior, namely the formation of reference point. In our model, if a candidate announce to be high valence, he induces his electorate to expect a high payoff; if voters find out that he cannot deliver this payoff (because his valence is lower than what initially claimed), a positive mass of voters may decide to vote for the candidate of the opposing party in the attempt of reducing the disappointment associated with low valence. If the mass of voters who change their electoral behavior after a lie represents a sufficiently high fraction of the population, this type of behavior will induce candidates to reveal their type sincerely. We further show that the range of parameters for which this truthful equilibrium exists is increasing in the degree of loss aversion.

The interaction between reference point and political announcements is an interesting topic for future research. On the one hand, it may be interesting understanding whether and under which conditions reference dependence may induce politicians to support false representations of reality and chain of lies even though this would not be *ex-ante* optimal. On the other hand, in this paper we only looked at politicians' statement concerning their own type, but future research should try to account for the possibility of statements concerning the valence of opponents. Indeed, in a model with reference-dependent preferences lowering the belief of voters concerning the quality of the opponent may be counterproductive if this person turns out to be better than anticipated.

3.6 Proofs

3.6.1 Proof of Proposition 19

The proof of this proposition is rather involved and we will split it in different steps. Throughout the proof we will maintain the assumption that voters assign probability 1 to the candidates playing a truthful communication strategy.

Reference utility associated with dynamic consistent strategies.

We start our analysis by characterizing the reference utility associated with dynamic consistent strategies

Lemma 2 Consider a candidate biased in favor of party r ($\gamma_i > 0$). Then the reference utilities associated with dynamic consistent strategies given $\left(t_i^{Tr}, w_i\right)_{i\in\{\ell,r\}}$ are described as follows:

i) if $\left(m^\ell, m^r\right) = (\theta_L, \theta_L)$, then $\tilde{u}\left(\rho; \gamma_i, (\theta_L, \theta_L), \chi\right) = \delta_\kappa \ \forall \gamma_i$,

ii) if $\left(m^\ell, m^r\right) = (\theta_H, \theta_H)$, then $\tilde{u}\left(\rho; \gamma_i, (\theta_H, \theta_H), \chi\right) = \delta_1 \ \forall \gamma_i$,

iii) if $\left(m^\ell, m^r\right) = (\theta_L, \theta_H)$, then $\tilde{u}\left(\rho; \gamma_i, (\theta_L, \theta_H), \chi\right) = \delta_1 \ \forall \gamma_i$,

iv) if $\left(m^\ell, m^r\right) = (\theta_H, \theta_L)$, then there exist $0 < \gamma^{**}(\kappa, \eta, \lambda) < \gamma^*(\kappa) < 1$ and $\alpha^*(\kappa, \eta, \lambda, \gamma)$ such that:

$$\tilde{u}\left(\rho; \gamma_i, (\theta_H, \theta_L), \chi\right) = \begin{cases} \delta_\kappa & \text{if } \gamma_i \in (\gamma^*(\kappa), 1] \\ \tilde{u} \in U & \text{if } \gamma_i \in [\gamma^{**}(\kappa, \eta, \lambda), \gamma^*(\kappa)] \\ (1-\gamma_i)[1] + \gamma_i[0] & \text{if } \gamma_i \in [0, \gamma^{**}(\kappa, \eta, \lambda)) \end{cases},$$

where:

$$U = \left\{ \begin{array}{c} \alpha^*[\kappa] + (1-\alpha^*)(1-\gamma_i)[1] + (1-\alpha^*)\gamma_i[0], \\ \delta_\kappa, \ \ (1-\gamma_i)[1] + \gamma_i[0] \end{array} \right\}$$

A symmetric characterization holds for candidates with bias $\gamma_i < 0$.

Proof: Consider a voter with bias $\gamma_i > 0$. We will analyze all pairs of announcements separately.

Suppose that $\iota_A = (\theta_L, \theta_L)$ in period 0. Then $\pi_1^{\ell}(\theta_L, \theta_L) = \pi_1^{r}(\theta_L, \theta_L) = 0$ and our assumption concerning beliefs updating implies that voters will keep assigning probability 0 to the candidate being high valence regardless of the actual signal received. Thus, the voters will behave in the same way regardless of the signal received and consequently the reference utility depends on a single action taken by the voters. Observe that the utility from voting for the candidate of the left party is given by $(1-\gamma_i)[\kappa] + \gamma_i[0]$, while the utility associated with voting for the candidate of the right party is given by κ. Suppose that the voter is thinking of voting for the candidate of the left party. Then his reference utility would be $(1-\gamma_i)[\kappa] + \gamma_i[0]$ and his utility from following this strategy would be $(1-\gamma_i)\kappa - (1-\gamma_i)\gamma_i\kappa\eta(\lambda-1)$. On the other hand if the candidate were to vote for the candidate party r, his utility would be $\kappa + \gamma_i\eta\kappa$. Clearly, the latter utility is greater than the former for any $\gamma_i \in (0,1)$. We conclude that voting for the candidate of the left party at information set (θ_L, θ_L) cannot be a dynamic consistent strategy. Now, suppose that the candidate thinks of voting for the candidate of the right party. Then his reference utility would be δ_κ and by voting for the candidate of the right party the candidate will expect to get a total utility equal to κ. On the other hand, if he were to vote for the candidate of party ℓ, his utility would be $(1-\gamma_i)\kappa - \gamma_i\kappa\eta\lambda$. Clearly the former utility is higher than the latter (as well as of than any mixture between these two utilities) for any $\gamma_i > 0$. Finally, assume that the voter thinks of randomizing between the two candidates with some probability $y \in (0,1)$; it is easy to show that none of these strategies satisfy dynamic consistency since, given the induced reference utility, the voter would prefer voting for the right candidate with certainty. Therefore, at information set (θ_L, θ_L), the reference utility associated with the unique dynamic consistent strategy will be given by a degenerate measure over κ, δ_κ.

An analogous reasoning shows that after announcements (θ_H, θ_H), the reference utility associated with the unique dynamic consistent strategy is given by δ_1.[50] Furthermore, it is also immediate to see that the same holds after announcements (θ_L, θ_H).

Finally, consider announcements (θ_H, θ_L). In this case the voter assign probability 1 only to information sets $\iota_s \in \{((\theta_H, 1), (\theta_L, 0)), ((\theta_H, 1), (\theta_L, 1))\}$ and in both cases he will behave in the same way (this is a consequence of the fact that we are assuming that the voter disregards the signals received by an agent who self declared as a low valence candidate). Thus the voter can decide to follow one of the following three strategies: voting with certainty for (what he believes to be) the high candidate of party ℓ, voting with certainty for (what he believes to be) the low candidate of party r or randomizing and voting for the left candidate with probability $(1-y)$ and for the right one with probability y. We will analyze all these cases separately.

Suppose that the voter is thinking of voting for the candidate of the left party. In this case the reference utility would be $(1-\gamma_i)[1] + \gamma_i[0]$ and his utility from following the strategy would be given by $(1-\gamma_i) - \eta\gamma_i(1-\gamma_i)(\lambda-1)$. If instead, he were to vote for the candidate of party r, his utility would be $\kappa + \eta(\gamma_i\kappa - (1-\gamma_i)(1-\kappa)\lambda)$. Observe that:

$$(1-\gamma_i) - \eta\gamma_i(1-\gamma_i)(\lambda-1) \geq \kappa + \eta(\gamma_i\kappa - (1-\gamma_i)(1-\kappa)\lambda)$$

$$\Longleftrightarrow$$

$$\gamma_i^2\eta(\lambda-1) + \gamma_i(\kappa\eta(\lambda-1) - (1-\eta) - 2\eta\lambda) + (1-\kappa)(1+\eta\lambda) \geq 0 \qquad (3.11)$$

Observe that the two roots of the polynomial on the left hand side of inequality (3.11) are given by $\left(1-\kappa, \frac{(\lambda\eta+1)}{(\lambda-1)\eta}\right)$. Since $\frac{(\lambda\eta+1)}{(\lambda-1)\eta} \geq 1$, we conclude that inequality (3.11) is satisfied as long as $\gamma_i \leq 1-\kappa$. Define $\gamma^*(\kappa) = 1-\kappa$. Thus for all voter with bias $\gamma_i < \gamma^*(\kappa)$, there exists

[50] In this case the characterization of the dynamic consistent strategy would require to specify the behavior of voters even after the out-of-equilibrium information set in which voters receive signal $s^i = 0$ for some i (this corresponds to a case in which the announcement of the candidate is falsified by the signal). However, $\chi(. \mid (\theta_H, \theta_H))$ assigns probability 0 to these information sets and consequently the determination of the reference utility is unaffected by these out-of-equilibrium actions.

a dynamic consistent strategy at information (θ_H, θ_L) whose associated reference utility is given by $(1-\gamma_i)[1]+\gamma_i[0]$.

Suppose instead that the voter is thinking of voting for the candidate of the right party. In this case his reference utility would be δ_κ.If the voter decides to vote for the candidate of party r, his utility would be r, while if he were to vote for the candidate of party ℓ would be $(1-\gamma_i)+\eta((1-\gamma_i)(1-\kappa)-\gamma_i\kappa\lambda)$. The former is greater than the latter as long as:

$$\gamma_i \geq \frac{(1-\kappa)(1+\eta)}{1+(1-\kappa)\eta+\kappa\eta\lambda} = \gamma^{**}(\kappa,\eta,\lambda)$$

Observe that $\frac{\partial\gamma^{**}(\kappa,\eta,\lambda)}{\partial\lambda} < 0$, $\gamma^{**}(\kappa,\eta,1) = 1-\kappa$ and $\lim_{\lambda\to\infty}\gamma^{**}(\kappa,\eta,\lambda) = 0$, so that for any $\lambda > 1$, $\gamma^{*}(\kappa) > \gamma^{**}(\kappa,\eta,\lambda) > 0$. Thus if $\gamma_i > \gamma^{**}(\kappa,\eta,\lambda)$, there exists a dynamic consistent strategy whose associated reference utility is given by δ_κ.

Finally, when $\gamma \in [\gamma^{**}(\kappa,\eta,\lambda), \gamma^{*}(\kappa,\eta,\lambda)]$, there exists a third dynamically consistent reference point associated with a mixed strategy. Finally, suppose that the voter with bias γ_i thinks of voting for the candidate of party ℓ with probability α and for the candidate of party r with complementary probability. In this case the reference utility would be given by:

$$\alpha[\kappa] + (1-\alpha)(1-\gamma_i)[1] + (1-\alpha)\gamma_i[0]$$

This can be a dynamically consistent reference point only if at the electoral stage the voter is indeed willing to randomize. Thus the following indifference condition has to be satisfied:

$$\begin{aligned}\kappa + \eta((1-\alpha)\gamma_i\kappa - (1-\alpha)(1-\gamma_i)(1-\kappa)\lambda) = \\ = (1-\gamma_i) + \eta\alpha((1-\gamma_i)(1-\kappa)-\gamma_i\kappa\lambda) + \\ + \eta(1-\alpha)\gamma_i(1-\gamma_i) - \eta(1-\alpha)\gamma_i(1-\gamma_i)\lambda\end{aligned}$$

or equivalently:

$$\alpha = \frac{(\kappa - (1-\gamma_i) + \eta(\kappa\gamma_i - \lambda(1-\kappa)(1-\gamma_i) + (\lambda-1)\gamma_i(1-\gamma_i)))}{\eta((\lambda-1)\gamma_i(1-\gamma_i) - (\lambda-1)\kappa\gamma_i - (\lambda-1)(1-\kappa)(1-\gamma_i))}$$

Define the right hand side of the previous equality $\alpha^*(\kappa, \gamma, \eta, \lambda)$. Observe that the denominator of this expression is negative. Thus

$$\frac{(\kappa - (1-\gamma_i) + \eta(\kappa\gamma_i - \lambda(1-\kappa)(1-\gamma_i) + (\lambda-1)\gamma_i(1-\gamma_i)))}{\eta((\lambda-1)\gamma_i(1-\gamma_i) - (\lambda-1)\kappa\gamma_i - (\lambda-1)(1-\kappa)(1-\gamma_i))} < 1$$

$$\Longleftrightarrow$$

$$\gamma_i > \gamma^{**}(\kappa, \eta, \lambda)$$

and

$$\frac{(\kappa - (1-\gamma_i) + \eta(\kappa\gamma_i - \lambda(1-\kappa)(1-\gamma_i) + (\lambda-1)\gamma_i(1-\gamma_i)))}{\eta((\lambda-1)\gamma_i(1-\gamma_i) - (\lambda-1)\kappa\gamma_i - (\lambda-1)(1-\kappa)(1-\gamma_i))} > 0$$

$$\Longleftrightarrow$$

$$(\kappa - (1-\gamma_i) + \eta(\kappa\gamma_i - \lambda(1-\kappa)(1-\gamma_i) + (\lambda-1)\gamma_i(1-\gamma_i))) < 0$$

The left hand side of this inequality defines a quadratic form in γ_i with two roots $1-\kappa$ and $\frac{(\lambda\eta+1)}{\lambda\eta-\eta}$. Since the larger root ($\frac{(\lambda\eta+1)}{\lambda\eta-\eta}$) is greater than 1, we conclude that this inequality will be satisfied as long as $\gamma_i < 1-\kappa$. Thus if $\gamma_i \in [\gamma^{**}(\kappa, \eta, \lambda), \gamma^*(\kappa, \eta, \lambda)]$, $\alpha^*(\kappa, \gamma, \eta, \lambda)$ takes an admissible value and we have a third dynamic consistent strategy whose associated reference utility is given by:

$$\alpha^*(\kappa, \gamma, \eta, \lambda)[\kappa] + (1 - \alpha^*(\kappa, \gamma, \eta, \lambda))(1-\gamma_i)[1] + (1 - \alpha^*(\kappa, \gamma, \eta, \lambda))(1-\gamma_i)[0]$$

This conclude the characterization of the reference utilities associated with dynamically consistent strategies for voters with positive bias. The analysis for voters with negative bias ($\gamma_i < 0$) is analogous and omitted. ∎

Lemma 2 characterizes the reference utilities associated with dynamic consistent strategies. As, we should expect, if a voter is strongly biased in favor of one of the parties ($|\gamma_i|$ high), the only dynamic consistent strategy is to vote according to the ideological bias and, in this case, the reference utility will be either δ_1 or δ_κ. On the opposite, if the bias is small in absolute value, a voter, who has to decide between a low valence candidate from the party he is biased toward and a high valence candidate from the opposing party, will prefer the latter and consequently the reference utility will be given by the lottery $(1-\gamma_i)[1]+\gamma_i[0]$ (in all other cases voters with low bias will vote for the candidate of their preferred party). Finally when the bias assume intermediate values, both strategies are dynamic consistent and a third one arises; in this third strategy the voter selects the candidate of his preferred party with probability $\alpha^*(\kappa,\gamma,\eta,\lambda)$.

Reference utility associated with the optimal dynamic consistent strategy.

Lemma 2 implies that for intermediate values of biases more than one strategy is dynamic consistent at information set (θ_H,θ_L) and (θ_L,θ_H). According to the optimality requirement built in Definition 3, the following proposition characterizes the reference utility associated with the optimal dynamic consistent strategy.

Lemma 3 Assume candidates play a fully revealing equilibrium and consider a voter with bias $\gamma_i \in [\gamma^{**}(\kappa,\eta,\lambda), \gamma^*(\kappa)]$. Then the reference utility associated with the optimal dynamic consistent strategy at (θ_H,θ_L) is δ_κ. If $\gamma_i \in [-\gamma^*(\kappa), -\gamma^{**}(\kappa,\eta,\lambda)]$, then the reference utility associated with the optimal dynamic consistent strategy at node (θ_L,θ_H) is δ_κ.

Proof: Take a voter with bias $\gamma_i \in [\gamma^{**}(\kappa,\eta,\lambda), \gamma^*(\kappa)]$. Then, the reference utility associated with each of the three possible dynamic consistent strategies can be summarized in the

following table:

Prob. of supporting party r	Reference Utility
1	κ
0	$(1-\gamma_i)[1]+\gamma_i[0]$
$\alpha^*(\kappa,\gamma_i,\eta,\lambda)$	$\alpha^*[\kappa]+(1-\alpha^*)(1-\gamma_i)[1]+(1-\alpha^*)(1-\gamma_i)[0]$

Similarly, the total expected utility associated with each dynamic consistent strategy can be summarized as follows:[51]

Prob. of supporting party r	Expected Total Utility
1	κ
0	$(1-\gamma_i)-(1-\gamma_i)\gamma_i\eta(\lambda-1)$
$\alpha^*(\kappa,\gamma_i,\eta,\lambda)$	$(\kappa+\eta((1-\alpha^*)\gamma_i\kappa-(1-\alpha^*)(1-\gamma_i)(1-\kappa)\lambda))$

Thus the total utility expected utility associate with voting for the candidate of party r is higher than the one from voting for the candidate of party ℓ as long as:

$$(\lambda-1)\eta\gamma_i^2-((\lambda-1)\eta+1)\gamma_i+1-\kappa<0 \tag{3.12}$$

Under the restriction of Assumption 9, the polynomial on the left hand of (3.12) has two real roots given by:

$$\frac{(\lambda-1)\eta+1\pm\sqrt{1+\eta^2+\eta^2\lambda^2-2\eta^2\lambda-2\eta\lambda+2\eta+4\eta\lambda\kappa-4\eta\kappa}}{2\eta(\lambda-1)}$$

It is easy to show that for any parameter values the highest root is greater than 1 and the lowest is greater than 0. Therefore (3.12) inequality is satisfied as long as $\gamma_i\in(\gamma^{***}(\kappa,\eta,\lambda),1)$,

[51] To compute the utility associated with the mixed strategy, we exploited the fact that $\alpha^*(\kappa,\eta,\lambda,\gamma_i)$ is constructed in order to leave the voter with bias γ_i indifferent between voting for the candidate of party r or the one of party ℓ.

where

$$\gamma^{***}(\kappa,\eta,\lambda) := \frac{(\lambda-1)\eta+1-\sqrt{1+\eta^2+\eta^2\lambda^2-2\eta^2\lambda-2\eta\lambda+2\eta+4\eta\lambda\kappa-4\eta\kappa}}{2\eta(\lambda-1)}$$

Note that for any values of parameters $\lim_{\lambda\to 1}\gamma^{***}(\kappa,\eta,\lambda)=1-\kappa$, and at the same time $\lim_{\lambda\to\infty}\gamma^{***}(\kappa,\eta,\lambda)=0$. Furthermore $\frac{\partial\gamma^{***}(\kappa,\eta,\lambda)}{\partial\lambda}<0$. We will now compare $\gamma^{***}(\kappa,\eta,\lambda)<\gamma^{**}(\kappa,\eta,\lambda)$. Note that:

$$\frac{(1-\kappa)(1+\eta)}{1+(1-\kappa)\eta+\kappa\eta\lambda}> \\ > \frac{(\lambda-1)\eta+1-\sqrt{1+\eta^2+\eta^2\lambda^2-2\eta^2\lambda-2\eta\lambda+2\eta+4\eta\lambda\kappa-4\eta\kappa}}{2\eta(\lambda-1)}$$

if and only if

$$(1+(1-\kappa)\eta+\kappa\eta\lambda)\,C(\eta,\lambda,\kappa)> \\ > ((\lambda-1)\eta+1)(1+(1-\kappa)\eta+\kappa\eta\lambda)-(1-\kappa)(1+\eta)2\eta(\lambda-1)$$

where $C(\eta,\lambda,\kappa)=\sqrt{1+\eta^2+\eta^2\lambda^2-2\eta^2\lambda-2\eta\lambda+2\eta+4\eta\lambda\kappa-4\eta\kappa}$. Note that the right hand side of the previous inequality is positive. Therefore this inequality is satisfied if

$$(1+(1-\kappa)\eta+\kappa\eta\lambda)^2\left(1+\eta^2+\eta^2\lambda^2-2\eta^2\lambda-2\eta\lambda+2\eta+4\eta\lambda\kappa-4\eta\kappa\right)> \\ > (((\lambda-1)\eta+1)(1+(1-\kappa)\eta+\kappa\eta\lambda)-(1-\kappa)(1+\eta)2\eta(\lambda-1))^2$$

or equivalently if

$$4\kappa\eta^3(1-\kappa)(\lambda-1)^2(\kappa+\lambda-\kappa\lambda+\lambda\eta)>0 \tag{3.13}$$

Since 3.13 holds, we can conclude that $\gamma^{***}(\kappa,\eta,\lambda)<\gamma^{**}(\kappa,\eta,\lambda)$. Therefore, as long as $\gamma_i\in[\gamma^{**}(\kappa,\eta,\lambda),\gamma^{*}(\kappa)]$, the utility from voting for the right candidate of the right party will exceed the one for voting for the candidate of the left party.

Now, let us compare the utility when the reference point is κ with the utility when the reference point is $\alpha^* [\kappa] + (1-\alpha^*)(1-\gamma_i)[1] + (1-\alpha^*)(1-\gamma_i)[0]$. The formers is greater than the latter as long:

$$\kappa \geq (\kappa + \eta((1-\alpha^*)\gamma\kappa - (1-\alpha^*)(1-\gamma)(1-\kappa)\lambda))$$

or equivalently

$$\lambda > \frac{\gamma\kappa}{(1-\gamma)(1-\kappa)} \tag{3.14}$$

Observe that the right hand side of the equation (3.14) is increasing in γ. Thus it is maximized at $\gamma = 1-\kappa$. At this value the condition is satisfied for any $\lambda \geq 1$. We conclude that for any $\gamma_i \in [\gamma^{**}(\kappa,\eta,\lambda), \gamma^*(\kappa)]$, the utility associated with the strategy that prescribes to vote for the candidate of the right party will always exceeds the one associated with the strategy that votes for him only with probability $\alpha^*(\kappa,\gamma_i,\eta,\lambda)$.

Thus if $\gamma_i \in [\gamma^{**}(\kappa,\eta,\lambda), \gamma^*(\kappa)]$, the optimal dynamic consistent strategy will prescribe to always vote for the candidate of party r; therefore the reference utility associated with the optimal dynamic consistent strategy will be given by δ_κ.

This conclude the analysis for the voter with positive biases. The analysis for voters with negative biases is similar and omitted. ∎

Characterization of the optimal dynamic consistent strategy

Lemmata 2 and 3 characterize the reference utility associated with the optimal dynamic consistent strategy for any possible pair of announcements under the assumption that candidates have announced their valence sincerely. The following table summarizes the reference utility associated with the optimal dynamic consistent strategy $\rho^*(.,.)$ of a voter with posi-

tive bias:

$$\tilde{u}\left(\rho^{*};\gamma_{i},\iota_{A},\chi\right)=\begin{cases}\delta_{1} & \text{if } \iota_{A}=\left(\theta_{H},\theta_{H}\right),\ \iota_{A}=\left(\theta_{L},\theta_{H}\right)\\ \delta_{\kappa} & \text{if } \iota_{A}=\left(\theta_{L},\theta_{L}\right) \text{ or } \iota_{A}=\left(\theta_{H},\theta_{L}\right)\\ & \text{and } \gamma_{i}>\left(\gamma^{**}\left(\kappa,\eta,\lambda\right),1\right]\\ \varphi\in\Phi & \text{if } \iota_{A}=\left(\theta_{H},\theta_{L}\right) \text{ and } \gamma_{i}=\gamma^{**}\left(\kappa,\eta,\lambda\right)\\ \left(1-\gamma_{i}\right)[1]+\gamma_{i}[0] & \text{if } \iota_{A}=\left(\theta_{H},\theta_{L}\right) \text{ and } \gamma_{i}\in\left[0,\gamma^{**}\left(\kappa,\eta,\lambda\right)\right)\end{cases}$$

where

$$\Phi=\left\{\delta_{\kappa},\ \left(1-\gamma_{i}\right)[1]+\gamma_{i}[0],\ \alpha^{*}[\kappa]+\left(1-\alpha^{*}\right)\left(1-\gamma_{i}\right)[1]+\left(1-\alpha^{*}\right)\left(1-\gamma_{i}\right)[0]\right\}.$$

The reference utility of a voter with a negative bias is defined analogously.

We are now ready to provide the full characterization of the optimal dynamic consistent strategy.

Consider a voter with bias $\gamma_{i}>0$. We will analyze his behavior for any information set $\iota_{S}\in\mathcal{I}_{S}$. First of all, consider an information set

$$\iota_{S}\in\left\{\left(\left(\theta_{H},1\right),\left(\theta_{H},1\right)\right),\left(\left(\theta_{L},0\right),\left(\theta_{L},0\right)\right),\left(\left(\theta_{L},0\right),\left(\theta_{H},1\right)\right)\right\}$$

In this cases, the definition of dynamic consistent strategy together with the characterization of reference utility we gave imply that $\rho\left(\gamma_{i},\iota_{S}\right)=1$.[52] To see why, consider, for example, information set $\iota_{S}=\left(\left(\theta_{H},1\right),\left(\theta_{H},1\right)\right)$. In this case all voters with positive bias have reference utility 1 and will assign probability 1 to both candidates being high valence. Therefore,

[52]Recall that we are looking only at voters with positive bias. The behavior of voters with negative bias is defined symmetrically.

whenever voters are not surprised by the signals, they will carry out with their strategies and vote for the candidate of party r with probability 1.

Then consider information set $((\theta_H, 1), (\theta_L, 0))$. Once more, the definition of the reference utility associated with the optimal dynamic consistent strategy implies:

$$\rho(\gamma_i, ((\theta_H, 1), (\theta_L, 0))) = \begin{cases} 0 & \text{if } \gamma_i < \gamma^{**}(\kappa, \eta, \lambda) \\ x \in [0, 1] & \text{if } \gamma_i = \gamma^{**}(\kappa, \eta, \lambda) \\ 1 & \text{otherwise} \end{cases}$$

Furthermore, our assumptions on belief updating imply that voters will behave in a similar way at information set $((\theta_H, 1), (\theta_L, 1))$.

Now consider information set $((\theta_L, 0), (\theta_H, 0))$. In this case the reference utility of a voter with positive bias will be δ_1, but $\pi_2^{\ell}((\theta_L, 0), (\theta_H, 0)) = \pi_2^{r}((\theta_L, 0), (\theta_H, 0)) = 0$. Thus, the voter will vote for the candidate of party r as long as:

$$\kappa + (\kappa - 1)\eta\lambda > (1 - \gamma_i)\kappa + (1 - \gamma_i)(\kappa - 1)\eta\lambda - \gamma_i\eta\lambda$$

which is always satisfied. We conclude that $\rho(\gamma_i, ((\theta_L, 0), (\theta_H, 0))) = 1 \ \forall\gamma_i$.

Consider instead information set $\iota_S = ((\theta_H, 1), (\theta_H, 0))$. In this case the reference point will be δ_1 and the beliefs at the moment in which voters vote will be $\pi_2^{\ell}((\theta_L, 0), (\theta_H, 0)) = 1$ and $\pi_2^{r}((\theta_L, 0), (\theta_H, 0)) = 0$. Thus a candidate with positive bias will vote for the candidate of party r only if:

$$\kappa + (\kappa - 1)\eta\lambda \geq (1 - \gamma_i) - \gamma_i\eta\lambda$$

or equivalently $\gamma_i \geq 1 - \kappa = \gamma^{*}(\kappa)$.

Thus:

$$\rho\left(\gamma_i, \left(\left(\theta_H, 1\right), \left(\theta_H, 0\right)\right)\right) = \begin{cases} 0 & \text{if } \gamma_i < \gamma^*\left(\kappa\right) \\ x \in [0,1] & \text{if } \gamma_i = \gamma^*\left(\kappa\right) \\ 1 & \text{if } \gamma_i > \gamma^*\left(\kappa\right) \end{cases}$$

Now consider $\iota_S = \left(\left(\theta_L, 0\right), \left(\theta_L, 1\right)\right)$. In this case the reference utility of the voters with positive bias is δ_κ and their beliefs are given by $\pi_2^\ell\left(\left(\theta_L, 0\right), \left(\theta_L, 1\right)\right) = \pi_2^r\left(\left(\theta_L, 0\right), \left(\theta_L, 1\right)\right) = 0$. Then it is easy to see that any voter biased in favor of party r will vote for the candidate of party r. Thus $\rho\left(\gamma_i, \left(\left(\theta_L, 0\right), \left(\theta_L, 1\right)\right)\right) = 1$.

Reasoning in a similar way, we can show that

$$\begin{aligned} \rho\left(\gamma_i, \left(\left(\theta_H, 0\right), \left(\theta_L, 0\right)\right)\right) &= \rho\left(\gamma_i, \left(\left(\theta_L, 1\right), \left(\theta_L, 0\right)\right)\right) = \\ &= \rho\left(\gamma_i, \left(\left(\theta_H, 0\right), \left(\theta_H, 1\right)\right)\right) = \\ &= \rho\left(\gamma_i, \left(\left(\theta_L, 1\right), \left(\theta_H, 1\right)\right)\right) = 1 \ \forall \gamma_i > 0 \end{aligned}$$

and that:

$$\begin{aligned} \rho\left(\gamma_i, \left(\left(\theta_H, 0\right), \left(\theta_H, 0\right)\right)\right) &= \rho\left(\gamma_i, \left(\left(\theta_L, 1\right), \left(\theta_L, 1\right)\right)\right) = \\ &= \rho\left(\gamma_i, \left(\left(\theta_H, 0\right), \left(\theta_L, 1\right)\right)\right) = \\ &= \rho\left(\gamma_i, \left(\left(\theta_L, 1\right), \left(\theta_H, 0\right)\right)\right) = 1 \ \forall \gamma_i > 0 \end{aligned}$$

Performing a similar analysis for the voters with negative bias, we conclude the proof of the proposition. ∎

3.6.2 Proof of Proposition 22

In an uninformative equilibrium, the only behavior of candidates that need to be described is the signal distortion chosen by the low valence candidates at the unique information set (m, m).

Note that since every information set $\iota_S \in \mathcal{I}_S(m, m)$ can be reached with positive probability as long as $w_\ell < 1$ and $w_r < 1$. Thus, as long as w_ℓ and w_r are lower than 1, dynamic consistency has to specify an optimal strategy at each information set ι_S. Since we are looking for a symmetric equilibrium, we will assume that $w_\ell = w_r = w$.

Let us consider the formation of reference utility first. We will characterize the behavior of voters with positive bias ($\gamma_i > 0$); the analysis for voters with negative biases is identical and omitted. We start observing that the reference utility of a voter will be a probability distribution over utilities 1, κ and 0.

Consider information set $\iota_S = ((m, 0), (m, 0))$ first. At this information set, voters assign probability 1 to both candidates being low valence. Voting for the candidate of party r is the strategy that maximizes the material payoff of voters with positive biases; furthermore this strategy also maximizes the psychological gain/loss component of utility: for any reference point whenever voting for party ℓ entails a gain, so does voting for r and whenever voting for r entails a loss so does voting for ℓ. We conclude that at information set $\iota_S = ((m, 0), (m, 0))$, a voter with positive bias will always vote for the candidate of the right party. An identical reasoning shows that the same is true at information sets $\iota_S = ((m, 0), (m, 1))$. Thus the only dynamic consistent strategy $\forall \gamma_i \geq 0$ will be $\rho(\gamma_i, ((m, 0), (m, 1))) = \rho(\gamma_i, ((m, 0), (m, 0))) = 1$.

Now consider information set $\iota_S = ((m, 1), (m, 1))$. In a symmetric equilibrium, both low valence candidates choose the same level of signal distortion and voters will assign probability $\frac{q}{q+(1-q)w}$ to each of the candidates being high valence. Thus, regardless of the actual reference point, the candidate of party r is associated with a higher expected utility, a higher expected psychological gain and a (possibly, weakly) lower expected loss. Thus in a dynamic consistent

strategy, every voter biased to the right will vote for the candidate of party r at information set $\iota_S = ((m,1),(m,1))$.

Consider information sets in $\{((m,0),(m,0)),((m,0),(m,1)),((m,1),(m,1))\}$. The reference utility of the voter with bias $\gamma_i > 0$ would be:

$$(1-q)^2(1-w)^2[\kappa] + (1-q)(1-w)q[1] + (1-q)(1-w)(1-q)w[\kappa] + \\ + ((1-q)w+q)q[1] + ((1-q)w+q)(1-q)w[\kappa]$$

or equivalently:

$$q[1] + (1-q)\left((1-q) - (1-2q)w + (1-q)w^2\right)[\kappa]$$

Finally, consider information set $\iota_S = ((m,1),(m,0))$. In this case the voter assign probability 0 (respectively, $\frac{q}{q+(1-q)w}$) to the candidate of the right (respectively, left) party having high valance. Consider a voter with bias γ_i and denote with α the probability with which the voter thinks of voting for the candidate of the right party, were information set $((m,1),(m,0))$ to happen.

We will analyze various case. Suppose the voter with bias γ_i thinks of voting with probability 1 for the candidate of the right party ($\alpha = 1$) at information set $((m,1),(m,0))$. In this case the reference utility of the voter would be: $q[1] + (1-q)[\kappa]$. Thus, the utility from carrying out the plan is given by:

$$\kappa - q(1-\kappa)\eta\lambda$$

If the candidate were thinking to vote with probability 1 for the candidate of party ℓ at information set $((m,1),(m,0))$ ($\alpha = 0$), his reference utility would be:

$$q\left(\left(1-q\right)\left(1-w\right)\left(1-\gamma_i\right)+1\right)\left[1\right]+$$
$$+\left(1-q\right)\left(qw+\left(1-q\right)\left(1-w\gamma_i+w^2\gamma_i\right)\right)\left[\kappa\right]+$$
$$+\left(\left(1-q\right)w+q\right)\left(1-q\right)\left(1-w\right)\gamma_i\left[0\right]$$

In this case the utility of a voter who carries out his initial plan would be:

$$\left(1-\gamma_i\right)\left(\frac{q+\left(1-q\right)w\kappa}{q+\left(1-q\right)w}\right)+\eta\left(1-q\right)\left(1-w\right)\left(\gamma_i-\gamma_i^2\right)\left(q+\left(1-q\right)w\kappa\right)+$$
$$+\eta\left(1-q\right)\left(qw+\left(1-q\right)\left(1-w\gamma_i+w^2\gamma_i\right)\right)\cdot$$
$$\cdot\left(\left(\left(1-\gamma_i\right)\frac{q}{q+\left(1-q\right)w}\right)\left(1-\kappa\right)-\gamma_i\kappa\lambda\right)+$$
$$-\eta\lambda q\left(\left(1-q\right)\left(1-w\right)\left(1-\gamma_i\right)+1\right)\left(\left(\left(1-\gamma_i\right)\frac{\left(1-q\right)w}{q+\left(1-q\right)w}\right)\left(1-\kappa\right)+\gamma_i\right)$$

Finally, if the voter were thinking to vote at information set $\left(\left(m,1\right),\left(m,0\right)\right)$ for the candidate of party r, with probability $\alpha\in\left(0,1\right)$, his reference utility would be the mixture of the previous two with weights α and $\left(1-\alpha\right)$, namely:

$$\left(\alpha q+\left(1-\alpha\right)q\left(\left(1-q\right)\left(1-w\right)\left(1-\gamma_i\right)+1\right)\right)\left[1\right]+$$
$$+\left(\alpha\left(1-q\right)+\left(1-\alpha\right)\left(1-q\right)\left(qw+\left(1-q\right)\left(1-w\gamma_i+w^2\gamma_i\right)\right)\right)\left[\kappa\right]+$$
$$+\left(1-\alpha\right)\left(\left(1-q\right)w+q\right)\left(1-q\right)\left(1-w\right)\gamma_i\left[0\right]$$

We will abuse notation and denote the reference utility associated γ_i and w with $\alpha\left(\gamma_i,w\right)$, or simply with α. If this were the reference point of the voter, his utility would be given by:

$$\kappa-q\left(1-\kappa\right)\eta\lambda-$$
$$-\eta\left(1-\alpha\right)\left(1-q\right)\left(1-w\right)\left(q\left(1-\gamma_i\right)\left(1-\kappa\right)\lambda--\left(\left(1-q\right)w+q\right)\gamma_i\kappa\right)$$

or equivalently by:

$$(1-\gamma_i)\left(\frac{q+(1-q)\,w\kappa}{q+(1-q)\,w}\right)+\eta\,(1-q)\,(1-\alpha)\,(1-w)\left(\gamma_i-\gamma_i^2\right)(q+(1-q)\,w\kappa)-$$
$$-\eta q\,(1+(1-\alpha)\,(1-q)\,(1-w)\,(1-\gamma_i))\left(\frac{(1-\gamma_i)\,(1-q)\,w}{q+(1-q)\,w}\,(1-\kappa)+\gamma_i\right)\lambda+$$
$$+\eta\,(1-q)\,(\alpha+(1-\alpha)\,(qw+(1-q)\,(1-w\,(1-w)\,\gamma_i)))\cdot$$
$$\cdot\left(\frac{q\,(1-\gamma_i)}{q+(1-q)\,w}\,(1-\kappa)-\gamma_i\kappa\lambda\right)+$$

The value of α will have to adjust in order to make the two previous expressions equal.

Now observe that when $\gamma_i=1$, the former expression is greater than the second for any value of α, while if $\gamma_i=0$ the opposite is true.

Furthermore, for $\gamma_i=1$, the utility associated with reference point $\alpha=1$ is dynamically consistent and it is higher than the one associated with reference point $\alpha<1$, while if $\gamma_i=0$, the only dynamically consistent reference point is $\alpha=0$. Since all the functions we are analyzing are continuous in γ_i, by continuity we can conclude that there exist two threshold levels $0<\underline{\gamma}\leq\overline{\gamma}<1$ such that in the optimal dynamic consistent strategy, all voters with bias in $(\overline{\gamma},1]$ will vote for the right candidate and all those with bias lower than $\left[0,\underline{\gamma}\right)$ would vote for the left candidate.

Observe that the mixed equilibrium (as well as the one in which $\alpha=0$) can not be supported for any voters with bias γ_i, where γ_i is such that:

$$\kappa\geq(1-\gamma_i)\left(\frac{q+(1-q)\,w\kappa}{q+(1-q)\,w}\right)$$

(in this case, it is easy to see that the optimal dynamic consistent strategy will be to vote for the candidate of party r).

Therefore the mixed strategy can be an optimal dynamic consistent strategy only for those voters for which $\kappa<(1-\gamma_i)\left(\frac{q+(1-q)w\kappa}{q+(1-q)w}\right)$. Furthermore in this case, one can show that given the linear structure of the gain-loss component (and of the material payoff), whenever

the mixed strategy is dynamic consistent either or both of the other strategies are dynamic consistent and that the mixed strategy delivers a lower utility than one of the two degenerate strategies. Thus the optimal dynamic consistent strategy will prescribe to vote for party r at information set $((m,1),(m,0))$ with probability either 0 or 1.

Then we conclude that the function $\rho(.;((m,1),(m,0))) : [0,1] \to [0,1]$ is bounded and has at most two discontinuities (this follows from the utility from voting for the right party being constant in γ_i and the one from voting for the left party being quadratic in γ_i).

Replicating the same steps for $\gamma_i < 0$, we can show that $S_\ell((m,s_\ell),(m,s_r))$ and $S_r((m,s_\ell),(m,s_r))$ are well defined for any pair s_ℓ and s_r. Now observe that $S_\ell((m,s_\ell),(m,s_r))$ and $S_r((m,s_\ell),(m,s_r))$ depends on w and are continuous in it (the discontinuity points in which the bias of a voter leads him to switch from supporting party r to supporting party ℓ vary continuously in w).

Now consider the low candidate of party r. Given the previous analysis, he will choose signal distortion level w to solve:

$$\begin{aligned}&\max_{w\in[0,1]} w\left[\left(q+\bar{w}(1-q)\right)S_r((m,1),(m,1))+(1-q)(1-\bar{w})S_r((m,0),(m,1))\right]+\\&+(1-w)\left[\left(q+\bar{w}(1-q)\right)S_r((m,1),(m,0))+(1-q)(1-\bar{w})S_r((m,0),(m,0))\right]-\\&-\frac{cw^2}{2}\end{aligned}$$

In particular, this function will have a unique maximizer for any $S_r((b,s_\ell),(b,s_r))$. Then we can apply Brouwer fixed point theorem to conclude that there exists a fixed point of the mapping that determine the optimal level of signal distortion and conclude that a symmetric equilibrium exists. ∎

Chapter 4

Full Bayesian Implementation with Hard Evidence

4.1 Implementation and Hard Evidence

The effect of incentives on the behavior of agents lies at the core of modern economics. The theory of implementation addresses this issue in its generality and investigates the circumstances under which a social planner can use incentives schemes (mechanisms) to induce agents to behave in ways that result in (implement) some specific outcome. Given its abstract formulation, implementation theory can be used to analyze a wide range of different situations: a parliament may try to induce the members of the government to reveal accurate data on some bill, tax authorities may design taxation scheme to limit false declarations by tax-payers, parents may want their children to reveal the status of completion of their homework.

In the usual formulation of the implementation problem, each participant to the mechanism is characterized by some private information (type) the planner would like to acquire, but he can claim to have any possible information (in the implementation terminology, he is free to mimic any other type). Thus, in order to attain social planner's goals the mechanism has

to satisfy an obvious incentive compatibility constraint: each agent must be willing to reveal his true type (behave truthfully), even if he could lie and send any other declaration.

In reality, however, agents are not always free in their deceptions. The environment in which they act imposes constraints to the declarations they can possibly make: governments cannot lie to the parliament regarding public data, IRS has access to evidence concerning tax-payers income, parents may look at children's exercise-books. In short, the existence of evidence restricts the set of declarations an agent can plausibly make and this may help the planner to achieve his goal.

In this paper we study the problem of full implementation under incomplete information and in the presence of evidence. In particular, we focus on the implementation of social choice functions when evidence is hard, that is it cannot be counterfeited and the planner cannot prevent agents from sending any piece of evidence they have. The former assumption can be seen as the limit case in which counterfeiting is costly, while the latter can be justified assuming that the social planner is subject to a third party (e.g., representatives of the judicial system) who protects agents' right to support their statement with any evidence they possess.

We begin by providing a necessary condition for full implementation of a social choice function in general environments where evidence is available: the EIC-EBM condition. Intuitively, this condition states that the planner can require agents to provide evidence so that (i) if everybody else is announcing his type truthfully, each agent is willing to be truthful as well, (ii) if some agents are being deceitful, at least one agent is willing to reveal the deception by playing the role of a whistle-blower.[1] These conditions generalize similar ones used by Jackson (1991) for Bayesian implementation without evidence. We also show that the EIC-EBM condition is sufficient for full implementation when there are 3 or more agents and the disagreement among agents is sufficiently strong (environments in which the conflict

[1]Note that, although the planner cannot prevent agents from bringing any evidence they possess, he can still require them to provide some specific evidence to support their cheap talk announcement.

of interests among agents is strong enough are called economic). Although we do not provide a full characterization of the set of implementable social choice functions in general environments, we still provide a sufficient conditions for full implementation by adding a no-veto condition to the EIC-EBM one. In the particular case in which there is no hard evidence, our definitions coincide to the ones provided by Jackson (1991) and his results apply.

The conditions for implementation that we propose in this paper relate the evidence structure with the social choice function the planner wants to implement and, in general, may be cumbersome to check. However, under a specific assumption concerning the evidence structure that has received a lot of attention in the literature and that is often satisfied in applications, we show that the task of checking these conditions significantly simplifies.

The chapter is organized as follows: in the remaining of the introduction, we provide an overview of the relevant literature. Section 2 introduces the model and summarizes some known results on partial implementation. In Section 3 we define the EIC-EBM condition and we show that it is necessary for full implementation. Section 4 provides sufficient conditions for full implementation in environments with at least 3 agents. We first focus on economic environments and we then generalize the analysis to arbitrary environments. Section 5 concludes by pointing out some directions for future research and by commenting on our results.

4.1.1 Related Literature

In his seminal paper on implementation theory, Maskin (1999) provides conditions under which the social planner can fully implement a social choice rule. Although, Maskin's paper focused on full Nash implementation under complete information, the key insights of his work have been subsequently extended to many different settings. Moore and Repullo (1988) and Moore and Repullo (1990) provide a full characterization of the implementation prob-

lem in Nash and Subgame Perfect Equilibrium respectively.[2] Jackson (1991) addresses the issue of implementation in Bayes-Nash equilibria in environments with incomplete information.[3] Abreu and Sen (1991) and Abreu and Matsushima (1991) and Abreu and Matsushima (1992b) introduce the concept of virtual implementation (that is, implementation with probability close to 1) and provide a very appealing mechanism that attains this goal.[4]

In particular, Jackson (1991) provides necessary and sufficient conditions for Bayesian implementation. Whereas Maskin (1999) assumes that all agents know the true state of nature, Jackson (1991) studies situations in which each agent is not certain about the true state, although he may have some information about it. In these context he shows that Incentive Compatibility (IC) and Bayesian Monotonicity (BM) are necessary and almost sufficient condition for implementation in general environments.[5]

Jackson (1991)'s framework has the feature that the planner and the agents have common knowledge of the environment in which they are operating. In a series of influential papers, Bergemann and Morris (2005), Bergemann and Morris (2009b) and Bergemann and Morris (2010) analyze the problem of static implementation in environments in which there is no common knowledge about the beliefs of agents;[6] they refer to this approach as robust implementation. Penta (2011) analyzes the issue of robust implementation in dynamic environments. Oury and Tercieux (2011) look at a similar issue by characterizing the set social choice functions that can be implemented for a certain belief structure and for types close (in terms of interim beliefs) to it.

More recently some papers have investigated the role that evidence can play in the implementation problem. Bull and Watson (2007) analyze partial Bayesian implementation with hard

[2]Concerning Nash implementation see also Benoit and Ok (2006) and Ben-Porath and Lipman (2011) who removed the assumption of no-veto power usually made in this literature.

[3]See also Palfrey, Srivastava, and Postlewaite (1993).

[4]An interesting disussion on virtual implementation can be found in Glazer and Rosenthal (1992) and Abreu and Matsushima (1992a).

[5]Jackson (1991) also shows that IC and BM are necessary and sufficient in economic environments, that is environments in which there is a sufficient degree of disagreement among the participants to the mechanism.

[6]On this topic see also Bergemann and Morris (2008a), Bergemann and Morris (2008b), Bergemann and Morris (2009a) and Bergemann, Morris, and Tercieux (2011).

evidence and provide conditions under which the implementation problem can be mapped into more standard implementation problems.[7] We summarize some of these results in Section 2. Our work differs from Bull and Watson (2007) because we focus on necessary and sufficient condition for full implementation. Ben-Porath and Lipman (2011) and Kartik and Tercieux (2011) analyze a complete information environment and study the role of evidence in Subgame Perfect and Nash implementation respectively. Our paper differs from theirs since we analyze environments with incomplete information.

In particular, Kartik and Tercieux (2011) is the paper most closely related to ours. They introduce Evidence Monotonicity (a natural extension of Maskin monotonicity to environments with evidence) and investigate how costly evidence production may affect the implementation problem. Our paper can be viewed as the extension of Jackson (1991)'s classic results on Bayesian implementation to environments with evidence and high cost of counterfeiting by using the insights of Kartik and Tercieux (2011). This extension is associated with some difficulties. First of all, with incomplete information, the beliefs of each agent concerning the information owned by the other participants play an important role in the implementation problem. This happens both because this information may affect the utility function of agents and because the behavior of an agent may depend on the information he has.[8] As a result, even if an agent is certain that the other participants are playing a non-mixed profile of strategies, he will still regard the outcome of the mechanism as a random variable and, consequently, the rewards and punishments built by the planner will have to incorporate this uncertainty. Furthermore, in a complete information setting all participants share the same information and, because of this, unilateral deviations of a single agent can be detected by comparing his own declaration with the ones made by other agents.[9] This is no longer possible with incomplete information: each agent may have some exclusive information that

[7] See also Bull and Watson (2004).

[8] To put it differently, we could have a setting with common values and some piece of information, although payoff irrelevant, may still be strategically relevant.

[9] Of course, this approach is possible only if the number of participants is greater or equal than 3.

the planner needs to extract and, whenever the evidence structure is unable to distinguish between two states, the social choice function must then satisfy an obvious truthtelling constraint (incentive compatibility). Finally, in line with Kartik and Tercieux (2011), we provide conditions that guarantee the detection and elimination of undesirable behavior thanks either to a reversal in some agents' (expected) utility or to the evidence structure available. However, differently from Kartik and Tercieux (2011), we need to take into account that agents may try to provide evidence in order to profit from a false declaration. Once more, this problem arises because agent i may be the only provider of some information necessary to attain the social planner's implementation goal; if this happens, the mechanism designer must rely on the information announced by agent i and this exposes the mechanism to a higher risk of manipulation. In order to overcome this problem, for every piece of evidence sent, e_i, the planner has to construct an outcome function that discourages the deviations of all those types of agent i which could possibly provide e_i.

4.2 The Model

Consider a society made by n agents who participate in a mechanism; the set of agents will be denoted by $N = \{1, 2, ..., n\}$. All relevant information concerning the environment are summarized in a state of nature θ. Thus, θ describes the preferences of agents, their information and any aspect of the environment that may be relevant for the social planner decision. The set of states of nature is denoted with Θ. For simplicity, we will assume that Θ is finite. In state θ, each agent $i \in N$ may be uncertain about the true state of the nature; to model this, we assume that:

$$\Theta := \Theta_1 \times \Theta_2 \times ... \times \Theta_n$$

where Θ_i represents the set of information available to player i.

Whenever, we consider a finite profile of sets $(X_1, ..., X_l)$, we use a standard notation and we denote with $X := \times_{i=1}^{l} X$ the Cartesian product of these sets and with $X_{-i} := \times_{j \neq i} X_j$ the Cartesian product of all set, but set X_i. A generic element of X, X_i and X_{-i} is denoted with x, x_i and x_{-i} respectively. When no confusion arises, we will write x_i to denote the projection of $x \in X$ on X_i.

Thus $\theta_i \in \Theta_i$ is the information available to agent i in each state (θ_i, θ_{-i}), $\theta_{-i} \in \Theta_{-i}$. An agent with information $\hat{\theta}_i$ assigns positive probability only to those state of nature that belong to the set $\left\{(\theta_i, \theta_{-i}) \in \Theta : \theta_i = \hat{\theta}_i\right\}$. We assume the existence of a common prior $\pi \in \Delta(\Theta)$ and we represent the uncertainty on Θ_{-i} of agent i when he has information $\hat{\theta}_i$ with a probability measure $\pi\left(. \mid \hat{\theta}_i\right) \in \Delta(\Theta_{-i})$.[10] For simplicity, we assume that the prior has full support: for every θ $\pi(\theta) > 0$.[11] The complete information case is a specific case of this framework, in which for every i, Θ_i is a singleton and $\pi(. \mid \theta_i)$ is a degenerate measure on some θ_{-i}. As mentioned in the introduction, the case of full implementation with complete information and hard evidence has been extensively studied by Kartik and Tercieux (2011). Departing from standard implementation problems, we allow agents to bring hard-evidence (i.e. non-falsifiable information) to support their cheap talk announcements concerning the state of nature. In particular, the set of evidence available to player i in state θ is denoted with E_i^{θ}. Since we are assuming that the information of agent i is captured by θ_i, we will assume that for every θ_i and θ_{-i}, $E_i^{(\theta_i, \theta_{-i})} = E_i^{\left(\theta_i, \theta'_{-i}\right)}$.[12] Given this property we will denote with $E_i^{\theta_i}$ the set of evidence available to an agent who has information θ_i.

Agents have preferences over a set A of social alternatives. These preferences are parametrized by the state of nature θ. Therefore player i's preferences in state θ are represented by a utility function $u_i(., \theta) : A \to \mathbb{R}$. Let $h : \Theta \to A$ be a function mapping states of nature

[10] The assumption of a common prior is made for simplicity and the analysis can be adapted to incorporate environments in which it does not hold.

[11] If we drop the full support assumption, the insights behind our results will still be valid, but the actual definitions and results would require a more involved notation. For this reason, we prefer to focus on the case of full support.

[12] This assumption is without loss of generality: the set of states of nature can always be enlarged so that it holds.

into alternatives. Assume agent i has information θ_i and that the common prior is π; then we can extend the utility of agent i to function h in the usual way:

$$U_i(h, \theta_i) = \sum_{\theta_{-i}} u_i(h(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi(\theta_{-i} \mid \theta_i)$$

The implementation problem arises from the desire of a social planner to implement a social choice function (scf) $f : \Theta \to A$ without knowing the actual state of nature.[13,14] Given this lack of information, the planner must rely on the announcements of the agents, but, obviously, agents may lie in order to modify the planner decision in their own interest. Formally, in order to implement scf $f(.)$ the planner builds a *mechanism*, that is a structure:

$$\Gamma = \langle N, g, (M_i)_{i \in N} \rangle$$

where M_i is a set of messages available to agent i and $g : M_1 \times ... \times M_n \to A$ is an outcome function mapping agents' messages into social outcomes.

We say that a mechanism Γ partially implements a social choice function $f(.)$ in equilibrium concept K if Γ admits a K-equilibrium σ, such that for every θ, $f(\theta) = g(\sigma(\theta))$. We say that Γ fully implements $f(.)$ if all K-equilibrium Γ are such that for every θ, $f(\theta) = g(\sigma(\theta))$. In this paper, we focus on full Bayesian implementation, that is full implementation in Bayesian equilibrium. Thus we require that for every θ, all Bayesian Nash equilibria $\sigma^* = (\sigma_i^* : \Theta_i \to M_i)_{i \in N}$ are such that $g(\sigma^*(\theta)) = f(\theta)$.

[13] We can think of $f(.)$ as representing the preferences of an actual planner or as representing the aggregate preferences of agents at some pre-informational stage.

[14] Although the extension to social choice correspondence is possible, it would not entail any conceptual gain. Thus, we will focus only on social choice functions.

Definition 10 Given a mechanism $\Gamma = \left\langle N, g, (M_i)_{i \in N} \right\rangle$, a Bayes-Nash equilibrium for this game is a profile of strategies $(\sigma_i^* : \Theta_i \to M_i)_{i \in N}$ such that for every i, θ_i and m_i :

$$\sum_{\theta_{-i}} u_i \left(g \left(\sigma_i^* (\theta_i), \sigma_{-i}^* (\theta_{-i}) \right), \theta_i, \theta_{-i} \right) \pi (\theta_{-i} \mid \theta_i) \geq$$

$$\geq \sum_{\theta_{-i}} u_i \left(g \left(m_i, \sigma_{-i}^* (\theta_{-i}) \right), \theta_i, \theta_{-i} \right) \pi (\theta_{-i} \mid \theta_i)$$

So far we have not said much on the actual structure of message spaces M_i. To incorporate the difference between cheap talk announcements and hard evidence, we distinguish between:

- the set C_i that represents a finite set of cheap talk message that could be sent in every state θ.
- the set E_i^θ represents the finite set of hard evidence available to agent i in state θ.

We will denote with $E_i = \bigcup_\theta E_i^\theta$ the set of all pieces of evidence available to player i in some state and for every $e_i \in E_i$, we will write $\Theta_i (e_i)$ for the set of i's private information compatible with e_i, namely $\Theta_i (e_i) = \left\{ \theta_i : e_i \in E_i^{\theta_i} \right\}$. The correspondence $\varphi : \theta \rightrightarrows E$ associates to each state θ the subsets $E^\theta \subseteq E$ of evidence available in state θ for each agent i. We will refer to the pair (E, φ) as to the *evidence structure* of a specific environment. We will further denote with $\varphi_i (\theta)$ the projection of $\varphi (\theta)$ on the i-th dimension. Observe that we made two implicit assumption concerning hard-evidence: the social planner cannot restrict the evidence agents can bring and evidence cannot be counterfeited. The first assumption can be justified assuming that the social planner is subject to a legal system that protects agents' right to bring all the evidence they want to support their claims. The second can be interpreted as stating that the cost of evidence falsification is extremely high.[15]

An environment is a structure:

$$\mathcal{E} = \left\langle N, \varphi, \left(\Theta_i, \left(E_i^\theta \right)_{\theta \in \Theta}, u_i \right)_{i \in N} \right\rangle$$

[15] In the complete information case, Kartik and Tercieux (2011) analyzes the implementation problem with and without these assumptions.

where all the elements have been previously defined. In particular, the evidence structure associated to a certain environment can take several forms. Below we describe a few evidence structure that have been studied in the literature:

- *no hard evidence*: evidence does never distinguish among states. Formally, for every i, $\varphi_i(\theta) \equiv E_i$. In this case, our framework would coincide with Jackson (1991)'s one and all his results would apply.

- *fully state-revealing evidence structure*: agents are compelled to bring evidence that fully reveals their own private information. Formally, for every θ, θ' and for every i $\varphi_i(\theta) \cap \varphi_i(\theta') = \emptyset$. In this case the social planner can implement any social choice function he wants since the evidence structure enables him to learn the state of nature. Observe that a specific feature of this evidence structure is that agent i does not have access to a piece of evidence that he could provide independently of the state (such as the piece of evidence "no evidence").

- *normal evidence structure*: agents can bring a piece of evidence that summarizes all the evidence available to them in different states. Formally, for every i and θ, there is some $\bar{e}_{\theta_i} \in \varphi_i(\theta)$ such that:

$$\bar{e}_{\theta_i} \in \varphi_i(\theta') \Longrightarrow \varphi_i(\theta) \subseteq \varphi_i(\theta')$$

  The case of normal evidence structure has received great attention in the literature. In this paper, we will show that this evidence structure plays a key role because it is often satisfied in practice and it significantly simplifies the implementation problem.

The following examples provide situations in which the previous evidence structure holds.

Example 6 (Fully Revealing Evidence Structure) Suppose that the US Department of Homeland Security decides whether to admit or reject travelers based on their nationality

and on the list of countries that they visited over the last 2 years. Assume for simplicity that each traveler arrives to the US with a valid, non-counterfeitable electronic passport and that each entry in a foreign country leaves an electronic stamp in the passport. Then, independently of the cheap talk made by travelers, the DHS officer can recover the actual countries they visited by looking at the information in the passport's microchip.

Example 7 (Normal Evidence Structure) Suppose a firm is made by n divisions and let Θ be the set of possible total revenues. Assume that the director of each division knows the total profit of his own division, θ_i, but he is uncertain about θ_{-i}. The CEO of a firm wants the directors to truthfully report their profits, but each manager would like to divert part of the profits for personal purposes. In this case we can think of $\bar{e}_{\theta_i}$ as the set of all invoices and documents available to some manager: the manager can always decide to share all the documentation he has with the CEO. However the manager of a very productive division may decide to reveal only some of the evidence and claim a lower profit. The CEO (that is, the social planner) can then try to set up a compensation scheme (a mechanism) to induce truthful revelation.

The problem of Bayesian implementation with evidence has been addressed by Bull and Watson (2007) and their analysis has focused on partial (or weak) implementation. For completeness, we summarize below some of their definitions and results.

Definition 11 Given an environment $\mathcal{E}$, a social choice function f is partially implementable with direct and truthful messages if there is a mechanism

$$\Gamma = \left\langle N, g, (C_i, E_i)_{i \in N} \right\rangle$$

such that for every i $C_i = \Theta_i$, the mechanism has an equilibrium $(\sigma_i^* : \Theta_i \to M_i)_{i \in N}$ with $g^*(\sigma^*(\theta)) = f(\theta)$ for every θ and for every player i and information θ_i, $\sigma_i^*(\theta_i) = (\theta_i, e_i)$ for some $e_i \in E_i^{\theta_i}$.

Definition 12 Given an environment $\mathcal{E}$ with normal evidence structure, a social choice function $f(.)$ is partially implementable with direct and truthful messages and maximal evidence production if there is a mechanism

$$\Gamma = \left\langle N, g, (C_i, E_i)_{i \in N} \right\rangle,$$

such that for every i $C_i = \Theta_i$, the mechanism has an equilibrium $(\sigma_i^* : \Theta_i \to M_i)_{i \in N}$ with $g(\sigma^*(\theta)) = f(\theta)$ for every θ and for every player i and information θ_i, $\sigma_i^*(\theta_i) = (\theta_i, \bar{e}_{\theta_i})$.

Thus a scf is partially implementable with direct and truthful messages if there is a mechanism in which implementation can be attained by having the participants in the mechanism reporting their own private information and providing some hard evidence. In the particular case of normal evidence, we can add the further requirement that in the direct and truthful mechanism people provide the maximal informative piece of evidence. The following Theorems, due to Bull and Watson (2007), generalize the revelation principle to environments where agents can provide hard evidence.[16]

Theorem 23 ***(Theorem 1 in Bull and Watson (2007))*** If a social choice function is partially implementable, then it is partially implementable with direct and truthful messages.

Theorem 24 ***(Theorem 2 in Bull and Watson (2007))*** Fix an environment $\mathcal{E}$ with normal evidence structure. Then, if a social choice function is partially implementable, it must be partially implementable with direct and truthful messages and maximal evidence production.

4.3 Necessary Condition

In this section we provide a necessary conditions for Bayesian Implementation with evidence production, the EIC-EBM condition. As it will become clear, this condition is based on

[16] On the revelation principle see for example Myerson (1982).

the merge of two separate conditions Evidence Incentive Compatibility (EIC) and Evidence Bayesian Monotonicity (EBM). As suggested by name, these definitions represent the extensions of Incentive Compatibility (IC) and Bayesian Monotonicity (BM) as defined by Jackson (1991) to environments where evidence provision is available.[17]

We start introducing the important concept of deception. A deception is a function $\alpha_i : \Theta_i \to \Theta_i$; a deception for player i represents the information agent i is pretending to have when his information is θ_i. In standard Bayesian implementation an agent is free to mimic any type he wants; however, this is no longer true once we introduce hard-evidence: obviously, an agent with information θ_i cannot credibly pretend to be type $\alpha_i(\theta_i)$ if $E_i^{\alpha_i(\theta_i)} \cap E_i^{\theta_i} = \emptyset$. Furthermore, even if $E_i^{\alpha_i(\theta_i)} \cap E_i^{\theta_i} \neq \emptyset$, the evidence provided by the agent to support his cheap talk announcement $\alpha_i(\theta_i)$ may not be the one required by the planner. This suggests a fundamental distinction between the messages sent by the agents and the interpretations the planner decides to give to these messages.[18] The definition of EIC is based on this intuition.

Definition 13 Given a scf f and an environment $\mathcal{E}$, $(\eta_i : \Theta_i \to E_i)_{i \in N}$ is a profile of incentive compatible evidence selections with respect to f if:

(i) for every i and θ_i, $\eta_i(\theta_i) \in E_i^{\theta_i}$;

(ii) for every i and θ_i, if $\eta_i(\theta_i') \in E_i^{\theta_i}$ for some θ_i', then

$$\sum_{\theta_{-i}} u_i\left(f\left(\theta_i, \theta_{-i}\right), \theta_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i\right) \geq \sum_{\theta_{-i}} u_i\left(f\left(\theta_i', \theta_{-i}\right), \theta_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i\right);$$

(iii) for every i and for every e_i, we can find a function $d_{e_i} : \Theta \to A$, such that for every $\hat{\theta}_i \in \Theta_i(e_i)$ and for every $\theta_i' \in \Theta_i$:

$$\sum_{\theta_{-i}} u_i\left(f\left(\hat{\theta}_i, \theta_{-i}\right), \hat{\theta}_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \hat{\theta}_i\right) \geq \sum_{\theta_{-i}} u_i\left(d_{e_i}\left(\theta_i', \theta_{-i}\right), \hat{\theta}_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \hat{\theta}_i\right).$$

[17] Analyzing the complete information case, Kartik and Tercieux (2011) introduce evidence monotonicity and show that this is the natural extension of Maskin monotonicity to environments in which hard evidence is available. Our results on Bayesian implementation resemble to theirs.

[18] This issue is discussed in details in Bull and Watson (2007).

The evidence selections mentioned in Definition 13 represent the way in which the planner interprets the message sent by agents: in order to interpret a certain message m_i as claiming that the state is θ_i, the planner requires agent i to provide piece of evidence $\eta_i(\theta_i)$. Then, the first condition in Definition 13 imposes an obvious compatibility between the evidence structure and the interpretation of the planner. The other two conditions, instead, have to do with the strategic behavior of agents. The second condition is a truth-telling constraint in which the range of possible deviations is limited by the evidence structure and by the interpretation η_i imposed by the planner. The third condition is more subtle. Although the planner builds a mechanism with the intention to interpret some messages in certain ways, there is nothing that prevents agents from sending inconsistent or unexpected messages. If this is indeed the case, condition (iii) provides a function that the planner can use to punish an agent who has sent a cheap message inconsistent with the evidence e_i he has provided. Since agent i is the only one with information θ_i, this punishment needs to work for any possible information $\hat{\theta}_i$ compatible with the evidence provision e_i. Observe that this condition is always satisfied when every agent has a uniformly worst outcome, that is an outcome that is at the bottom of its ranking regardless of the actual state of nature. In this case, for every e_i, the function d_{e_i} can be defined as the one which always delivers the worst outcome regardless of the behavior of other players.[19] Evidence Incentive Compatibility (EIC) is closely related to Definition 13.

Definition 14 Given an environment $\mathcal{E}$, a social choice rule f satisfies Evidence Incentive Compatibility (EIC) if there exists a profile of incentive compatible evidence selections with respect to f.

[19]This problem and its solution are similar to those analyzed in a working paper version of Kartik and Tercieux (2011) in the case of implementation with one-agent: however in our context the condition is more involved because it has to take into account the private information of other agents as well.

Observe that without hard evidence EIC collapses to Jackson (1991)'s IC requirement.[20] On the opposite, if we assume a fully state-revealing evidence structure any social choice function satisfies EIC since the planner can disregard the cheap announcement of the agents and learn the state from the evidence they produce.

The next two examples show that the existence of hard evidence can help implementing certain functions and highlight the role played by condition (iii).

Example 8 Consider an environment with a unique agent and suppose there are two alternatives, $A = \{a, b\}$ and two states of nature $\Theta = \{\theta, \theta'\}$. Assume the agent knows the state so $(\Theta = \Theta_1)$ and that he prefers a to b regardless of the state. Consider the social choice function $f(\theta) = a$, $f(\theta') = b$. It is easy to see that without hard evidence this function cannot be implemented. Indeed this social choice function is not incentive compatible: in order to implement a in state θ, the mechanism must prescribe outcome a after a certain message m, but then the agent would like to send the same message also in state θ'. However, suppose we add the following (normal) evidence structure $E = \{e, e'\}$,

$$\varphi\left(\hat{\theta}\right) = \begin{cases} \{e, e'\} & \text{if } \hat{\theta} = \theta \\ \\ \{e'\} & \text{if } \hat{\theta} = \theta' \end{cases}$$

and suppose the planner picks the following evidence selection

$$\eta\left(\hat{\theta}\right) = \begin{cases} e & \text{if } \hat{\theta} = \theta \\ \\ e' & \text{if } \hat{\theta} = \theta' \end{cases}.$$

It is immediate to check that this is indeed an evidence compatible evidence selection once we define the function $d_e(\theta') = d_{e'}(\theta) = b$. Then the planner can fully implement the social

[20] In particular, without hard evidence the third condition in the definiton of an incentive compatible evidence structure is automatically implied by the second one.

choice function by specifying an outcome function that delivers a whenever the message is (θ, e) and b otherwise. To see this, observe that type θ has an obvious incentive to send message (θ, e), while type θ' is forced to provide evidence e', at which point his actual cheap message does not matter.

Example 9 Consider the following example borrowed by Kartik and Tercieux (2011). Let $A = \{a, b\}$ and $\Theta = \Theta_1 = \{\theta', \theta''\}$ and assume that the agent's preferences are given by (the numbers in the table represent the vNM utility indexes):

	θ	θ'
a	0	1
b	1	0

Suppose that the evidence structure is such that: $E_1^{\theta} = \{e', e\}$, $E_1^{\theta'} = \{e'', e\}$ and that the planner wants to implement social choice function $f(\theta) = a$, $f(\theta') = b$. Consider evidence selection $\eta_1(\theta') = e'$ and $\eta_1(\theta'') = e''$. Clearly this selection satisfies conditions (i) and (ii) in the definition of incentive compatible evidence selection. However it does satisfy condition (iii). The reason is obvious: no matter what the planner chooses after evidence e, one of the two agents will want to deviate and announce e in order to get his most preferred outcome. The reasoning for other evidence selections is similar, but they do not even satisfy condition (ii). Suppose now that we add a third alternative c so that the new preferences are given by:

	θ_1'	θ_1''
a	0	1
b	1	0
c	-1	-1

Now the planner can punish inconsistent messages with alternative c. Condition (iii) is satisfied. It is easy to see that in this case, we can construct a mechanism that implement the social choice function.

The next proposition states the intuitive result that EIC is a necessary condition for partial and, consequently, full implementation. As it will become clear with the proofs of theorems of Section 4.4, EIC is also sufficient for partial implementation.

Proposition 25 A scf $f(.)$ is partially implementable only if it satisfies EIC.

Proof: Suppose a scf is partially implementable in Bayes-Nash equilibrium, that is suppose there exists a mechanism $\Gamma = \left\langle N, g, (\Theta_i, E_i, u_i)_{i \in N} \right\rangle$ and a Bayes-Nash equilibrium σ^* of this mechanism such that for every θ, $g(\sigma^*(\theta)) = f(\theta)$. Then, Theorem 23 implies that the scf can be implemented with a direct mechanism. Then, we can find an equilibrium σ^* of the direct mechanism such that for every θ, $g(\sigma^*(\theta)) = f(\theta)$ and for every i, θ_i and $m_i \in \Theta_i \times E_i^{\theta}$:

$$\sum_{\theta_{-i}} u_i \left(g\left(\sigma_i^*(\theta_i), \sigma_{-i}^*(\theta_{-i}) \right), \theta_i, \theta_{-i} \right) \pi(\theta_{-i} \mid \theta_i) \geq$$

$$\geq \sum_{\theta_{-i}} u_i \left(g\left(m_i, \sigma_{-i}^*(\theta_{-i}) \right), \theta_i, \theta_{-i} \right) \pi(\theta_{-i} \mid \theta_i)$$

Define evidence selection $(\eta_i)_{i \in N}$ so that for every i and θ_i, $\eta_i(\theta_i) = proj_{E_i}(\sigma_i^*(\theta_i))$.
By construction $\eta_i(\theta_i) \in E_i^{\theta}$. Moreover, since for every θ, $g(\sigma^*(\theta)) = f(\theta)$, the previous inequality implies that for every θ_i' such that $\eta_i(\theta_i') \in E_i^{\theta_i}$:

$$\sum_{\theta_{-i}} u_i(f(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi(\theta_{-i} \mid \theta_i) \geq \sum_{\theta_{-i}} u_i(f(\theta_i', \theta_{-i}), \theta_i, \theta_{-i}) \pi(\theta_{-i} \mid \theta_i)$$

Finally consider any message e_i. For every θ_i and $\theta_{-i} \in \Theta_{-i}$ let:

$$d_{e_i}(\theta_i, \theta_{-i}) = g\left((\theta_i, e_i), \sigma_{-i}^*(\theta_{-i}) \right)$$

By the first inequality we conclude that for every i and e_i we can find a function $d_{e_i} : \Theta \to A$ such that for every $\hat{\theta}_i \in \Theta(e_i)$ and for every $\theta'_i \in \Theta_i$:

$$\sum_{\theta_{-i}} u_i\left(f\left(\hat{\theta}_i, \theta_{-i}\right), \hat{\theta}_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \hat{\theta}_i\right) \geq \sum_{\theta_{-i}} u_i\left(d_{e_i}\left(\theta'_i, \theta_{-i}\right), \hat{\theta}_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \hat{\theta}_i\right)$$

Thus $(\eta_i(.))_{i \in N}$ is a profile of incentive compatible evidence selections and, consequently, EIC is satisfied.

Proposition 25 provides a necessary condition for partial implementation. However EIC alone does not rule out the existence of other, undesirable, equilibria. The next definition introduce the concept of Evidence Bayesian Monotonicity (EBM) and will enable us to eliminate these other equilibria.

Definition 15 Given a scf f and an environment $\mathcal{E}$, we say that $(\eta_i : \Theta_i \to E_i)_{i \in N}$ is a profile of evidence Bayesian monotonic selections with respect to f if:
(i) for every i and θ_i, $\eta_i(\theta_i) \in E_i^{\theta}$
(ii) whenever a profile of deceptions α is such that for every i and for every $(\theta_1, ..., \theta_n)$, $\eta_i(\alpha_i(\theta_i)) \in E_i^{\theta_i}$ and $f(\alpha(.)) \neq f(.)$, then there exists an agent i, an information θ_i^* and a function $q : \Theta \to A$ such that

$$\sum_{\theta_{-i}} u_i\left(q\left(\alpha_i\left(\theta_i^*\right), \alpha_{-i}\left(\theta_{-i}\right)\right), \theta_i^*, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i^*\right) > \\ > \sum_{\theta_{-i}} u_i\left(f\left(\alpha_i\left(\theta_i^*\right), \alpha_{-i}\left(\theta_{-i}\right)\right), \theta_i^*, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i^*\right) \qquad (4.1)$$

while for every θ_i such that $\eta_i(\alpha_i(\theta_i^*)) \in E_i^{\theta_i}$:

$$\sum_{\theta_{-i}} u_i\left(f\left(\theta_i, \theta_{-i}\right), \theta_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i\right) \geq \sum_{\theta_{-i}} u_i\left(q\left(\alpha_i\left(\theta_i^*\right), \theta_{-i}\right), \theta_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i\right) \qquad (4.2)$$

Once more the first condition represents a trivial admissibility requirement. The second condition is the one that enables us to rule out undesirable equilibria. The intuition is as follows:

suppose agents are playing according to a deception profile $\alpha(.)$ and that this deception, in equilibrium, generates an outcome different from the one prescribed by $f(.)$. Assume that the social planner cannot find an evidence selection that detect this deception. Then, to destroy this bad equilibrium there must exists an agent i who, given a certain information, is willing to call the bluff. In order for this to happen, the planner must compensate this whistle-blower: the function q represents this compensation scheme and inequality (4.1) states that calling the bluff is indeed profitable. Inequality (4.2), on the other hand, states that the informer does not gain from falsely calling a bluff when everybody is being truthful. Once more, the definition of Evidence Bayesian Monotonicity is related to the existence of a profile of Bayesian monotonic evidence selections.

Definition 16 A social choice function f satisfies Evidence Bayesian Monotonicity (EBM) if we can find a profile of evidence Bayesian monotonic selections with respect to f.

In the special case of no hard evidence, the definition of Evidence Bayesian Monotonicity collapses to the Bayesian Monotonicity proposed by Jackson (1991). As already mentioned, EBM and EIC are both necessary to attain full implementation. The following example shows that an evidence selection can be incentive compatible, but not Bayesian monotonic.

Example 10 Consider the following environment $N = \{1, 2\}$, $A = \{a, b, c\}$, $\Theta_1 = \{\theta_1', \theta_1''\}$, $\Theta_2 = \{\theta_2', \theta_2''\}$, $E_1^{\theta'} = \{e_1', e_1''\}$, $E_1^{\theta''} = \{e_1''\}$, $E_2 = \{\bar{e}_2\}$. Let the preference of agent 1 be represented by the following vNM utility indexes:

	(θ_1', θ_2')	(θ_1', θ_2'')	(θ_1'', θ_2')	(θ_1'', θ_2'')
a	2	-2	1	1
b	3	-2	1	1
c	0	-10	0	0

and assume that agent 2 is indifferent between all outcome. Suppose that the common prior on the states is the following:

	θ_2'	θ_2''
θ_1'	$\frac{1}{3}$	$\frac{1}{6}$
θ_1''	$\frac{1}{6}$	$\frac{1}{3}$

and that the social planner wants to implement the following social choice function:

$$f(\theta) = \begin{cases} a & if\ \theta \in \{(\theta_1', \theta_2'), (\theta_1', \theta_2'')\} \\ b & if\ \theta \in \{(\theta_1'', \theta_2')\} \\ c & if\ \theta \in \{(\theta_1'', \theta_2'')\} \end{cases}$$

Two evidence selections are possible: (i) $\eta_1(\theta_1') = e_1'$, $\eta_1(\theta_1'') = e_1''$, (ii) $\eta_1(\theta_1') = e_1''$, $\eta_1(\theta_1'') = e_1''$. Consider first the selection $\eta_1(\theta_1') = e_1'$, $\eta_1(\theta_1'') = e_1''$. It is easy to see that this selection is incentive compatible (observes that in this case c is a uniformly worst outcome and so the third part of EIC definition is trivially satisfied). However this selection is not Bayesian monotonic. To see why, suppose player 2 is playing deception $\alpha_2(\theta_2') = \alpha_2(\theta_2'') = \theta_2'$, that is he is always announcing state θ_2' (since he is indifferent this is trivially optimal for him). Then, depending on player 1 announcement, only outcomes a or b will be implemented. Assume that agent 1 is playing deception $\alpha_1(\theta_1') = \alpha_1(\theta_1'') = \theta_1''$, that is he always pretends his information is θ_1''. Given the deception of the players outcome b will always be implemented. But note that for both types of player 1, b is the best alternative regardless of the type of his opponent. Therefore this selection is not Bayesian monotonic with respect to f and, as a consequence, undesirable equilibria are indeed possible. For completeness, we also point out that the alternative selection: $\eta_1(\theta_1') = e_1''$, $\eta_1(\theta_1'') = e_1''$ is not incentive compatible. To

see this, observe that type θ_1'' will have an incentive to announce θ_1' in order to get outcome a, regardless of the state.

Observe that the definitions of EIC and EBM requires the existence of two *distinct* profiles of evidence selection $(\eta_i(.))_{i\in N}$. Thus, in principle, the incentive compatible selection can differ from the Bayesian monotonic one, reflecting two different interpretations that the planner can give to the evidence received. Full implementation requires these two selections to coincide.

Definition 17 Given environment $\mathcal{E}$, a scf f satisfies the EIC-EBM condition if it admits a profile of evidence selections that is both incentive compatible and Bayesian monotonic with respect to f.

The next proposition states the main result of this section: the EIC-EBM condition is necessary to attain full implementation.

Proposition 26 Given an environment $\mathcal{E}$, a scf f is fully implementable only if it satisfies the EIC-EBM condition.

Proof: Suppose we can find a mechanism Γ that fully implements scf f. Then for any Bayes-Nash equilibrium of Γ, σ^*, we have that for every θ, $g(\sigma^*(\theta)) = f(\theta)$. Moreover, by Proposition 25 we know that the scf must satisfy EIC. Suppose there is no incentive compatible evidence selection with respect to f that is also Bayesian monotonic with respect to it. Then for any incentive compatible selection with respect to f, we can find a profile of deceptions $\alpha(.)$ such that:
(i) $f(\alpha(\theta)) \neq f(\theta)$ for some θ;
(ii) for every i and for every θ_i, $\eta_i(\alpha_i(\theta_i)) \in E_i^{\theta_i}$;

(iii) either for every agent i, every information θ_i and every function $q : \Theta \to A$:

$$\sum_{\theta_{-i}} u_i \left(f \left(\alpha_i \left(\theta_i \right), \alpha_{-i} \left(\theta_{-i} \right) \right), \theta_i, \theta_{-i} \right) \pi \left(\theta_{-i} \mid \theta_i \right) \geq$$

$$\geq \sum_{\theta_{-i}} u_i \left(q \left(\alpha_i \left(\theta_i \right), \alpha_{-i} \left(\theta_{-i} \right) \right), \theta_i, \theta_{-i} \right) \pi \left(\theta_{-i} \mid \theta_i \right)$$

or whenever we can find an agent i, an information θ_i^* and a function $q : \Theta \to A$ such that

$$\sum_{\theta_{-i}} u_i \left(q \left(\alpha_i \left(\theta_i^* \right), \alpha_{-i} \left(\theta_{-i} \right) \right), \theta_i^*, \theta_{-i} \right) \pi \left(\theta_{-i} \mid \theta_i^* \right) >$$

$$> \sum_{\theta_{-i}} u_i \left(f \left(\alpha_i \left(\theta_i^* \right), \alpha_{-i} \left(\theta_{-i} \right) \right), \theta_i^*, \theta_{-i} \right) \pi \left(\theta_{-i} \mid \theta_i^* \right),$$

we also have an information θ_i^{**} such that $\eta_i \left(\alpha_i \left(\theta_i^* \right) \right) \in E_i^{\theta_i^{**}}$ and

$$\sum_{\theta_{-i}} u_i \left(q \left(\alpha_i \left(\theta_i^* \right), \theta_{-i} \right), \theta_i^{**}, \theta_{-i} \right) \pi \left(\theta_{-i} \mid \theta_i^{**} \right) > \sum_{\theta_{-i}} u_i \left(f \left(\theta_i^{**}, \theta_{-i} \right), \theta_i^{**}, \theta_{-i} \right) \pi \left(\theta_{-i} \mid \theta_i^{**} \right).$$

Consider first the case in which for every agent i, every information θ_i and every function $q : \Theta \to A$:

$$\sum_{\theta_{-i}} u_i \left(f \left(\alpha_i \left(\theta_i \right), \alpha_{-i} \left(\theta_{-i} \right) \right), \left(\theta_i, \theta_{-i} \right) \right) \pi \left(\theta_{-i} \mid \theta_i' \right) \geq$$

$$\geq \sum_{\theta_{-i}} u_i \left(q \left(\alpha_i \left(\theta_i \right), \alpha_{-i} \left(\theta_{-i} \right) \right), \left(\theta_i, \theta_{-i} \right) \right) \pi \left(\theta_{-i} \mid \theta_i \right)$$

Fix a Bayes-Nash equilibrium σ^* of the mechanism Γ. By the assumption of full implementation, $\sigma^* \left(\theta \right) = f \left(\theta \right)$ for every θ. Build the strategy profile $\left(\hat{\sigma}_i : \Theta_i \to \Theta_i \times E_i \right)_{i \in N}$ such that for every i and θ_i $\hat{\sigma}_i \left(\theta_i \right) = \sigma_i^* \left(\alpha_i \left(\theta_i \right) \right)$. By construction this is an equilibrium and there is some state $\hat{\theta}$ such that $g \left(\hat{\sigma} \left(\hat{\theta} \right) \right) = g \left(\sigma^* \left(\alpha \left(\hat{\theta} \right) \right) \right) = f \left(\alpha \left(\hat{\theta} \right) \right) \neq f \left(\hat{\theta} \right)$. This contradicts the assumption of full implementation.

Suppose instead that for any agent i for which we can find an information θ_i^* and a function

$q : \Theta \to A$ such that:

$$\sum_{\theta_{-i}} u_i \left(q \left(\alpha_i \left(\theta_i^* \right), \alpha_{-i} \left(\theta_{-i} \right) \right), \theta_i^*, \theta_{-i} \right) \pi \left(\theta_{-i} \mid \theta_i^* \right) > \\ > \sum_{\theta_{-i}} u_i \left(f \left(\alpha_i \left(\theta_i^* \right), \alpha_{-i} \left(\theta_{-i} \right) \right), \theta_i^*, \theta_{-i} \right) \pi \left(\theta_{-i} \mid \theta_i^* \right),$$

we also have a state θ_i^{**} such that $\eta_i \left(\alpha_i \left(\theta_i^* \right) \right) \in E_i^{\theta_i^{**}}$ and

$$\sum_{\theta_{-i}} u_i \left(q \left(\alpha_i \left(\theta_i^* \right), \theta_{-i} \right), \theta_i^{**}, \theta_{-i} \right) \pi \left(\theta_{-i} \mid \theta_i^{**} \right) > \sum_{\theta_{-i}} u_i \left(f \left(\theta_i^{**}, \theta_{-i} \right), \theta_i^{**}, \theta_{-i} \right) \pi \left(\theta_{-i} \mid \theta_i^{**} \right). \tag{4.3}$$

Fix any Bayes-Nash equilibrium σ^* of the mechanism Γ. Since Γ fully implements f, $\sigma^* \left(\theta \right) = f \left(\theta \right)$ for every θ. Then either function q is not played after any message of agent i, in which case, we can replicate the previous reasoning and show that we cannot attain full implementation, or agent i has at least a message m_i^* such that function q is played when he sends this message. In this latter case, consider the strategy profile $\left(\hat{\sigma}_i : \Theta_i \to M_i \right)_{i \in N}$ such that for every $j \neq i$ and for every θ_j $\hat{\sigma}_j \left(\theta_j \right) = \sigma_j^* \left(\theta_j \right)$ and

$$\hat{\sigma}_i \left(\theta_i \right) = \begin{cases} \sigma_i^* \left(\theta_i \right) & \text{if } \theta_i \neq \theta_i^{**} \\ \\ m_i^* & \text{if } \theta_i = \theta_i^{**} \end{cases}$$

By construction, this strategy is a profitable deviation for agent i contradicting the assumption that σ^* is an equilibrium. Since this reasoning hold for any equilibrium we conclude that the hypothesis of full implementation is contradicted.

We conclude this section by pointing out a specific property of normal evidence structure. By Proposition 26, we know that to prove that a certain scf *cannot* be implemented, one would need to check any possible evidence selection. In environments with many pieces of evidence available, this could represent a quite challenging task. The existence of a

normal evidence structures significantly simplifies this task: the next proposition states that with normal evidence structure a scf f satisfies the EIC-EBM condition if and only if the evidence selection $(\eta_i^*)_{i\in N}$ in which for every i and for every θ_i, $\eta_i^*(\theta_i) = \bar{e}_{\theta_i}$ is both incentive compatible and Bayesian monotonic with respect to f. We will refer to this evidence selection as the maximally revealing evidence selection.

Proposition 27 Consider an environment $\mathcal{E}$ with normal evidence structure. Then a scf f satisfies the EIC-EBM condition if and only if the maximal revealing evidence selection is both incentive compatible and Bayesian monotonic.

Proof: The if part is trivially true. Therefore we will focus on the only if part. Suppose that we can find an evidence selection $(\eta_i)_{i\in N}$ satisfies incentive compatibility and Bayesian monotonicity. We will show that the maximally revealing evidence selection also does so. Clearly for every θ_i, $\eta_i^*(\theta_i) \in E_i^{\theta_i}$. Incentive compatibility requires that for every i and for every θ_i, $\eta_i(\theta_i') \in E_i^{\theta_i}$ implies

$$\sum_{\theta_{-i}} u_i\left(f\left(\theta_i,\theta_{-i}\right),\theta_i,\theta_{-i}\right)\pi\left(\theta_{-i}\mid\theta_i\right) \geq \sum_{\theta_{-i}} u_i\left(f\left(\theta_i',\theta_{-i}\right),\theta_i,\theta_{-i}\right)\pi\left(\theta_{-i}\mid\theta_i\right). \tag{4.4}$$

Then we can have two cases: $\bar{e}_{\theta_i'} \notin E_i^{\theta_i}$ or θ_i for which $\bar{e}_{\theta_i'} \in E_i^{\theta_i}$. In the former case, point (ii) in the definition of an incentive compatible evidence selection does impose any requirement. In the latter case, we can use the definition of normal evidence structure to conclude that $e_{\theta_i'} \in E_i^{\theta_i}$ implies $E_i^{\theta_i'} \subseteq E_i^{\theta_i}$. Therefore we can conclude that for every i and for every θ_i, $\eta_i^*(\theta_i') \in E_i^{\theta_i}$ for some θ_i' implies

$$\sum_{\theta_{-i}} u_i\left(f\left(\theta_i,\theta_{-i}\right),\theta_i,\theta_{-i}\right)\pi\left(\theta_{-i}\mid\theta_i\right) \geq \sum_{\theta_{-i}} u_i\left(f\left(\theta_i',\theta_{-i}\right),\theta_i,\theta_{-i}\right)\pi\left(\theta_{-i}\mid\theta_i\right).$$

Now consider condition (iii) in the definition of incentive compatible evidence selection. For every e_i, let d_{e_i} represent the function whose existence is guaranteed by condition (iii) in the definition of the incentive compatibility of evidence selection $(\eta_i)_{i\in N}$. Consider evidence

selection $(\eta_i^*)_{i\in N}$ and for every i, θ and e_i define function:

$$d_{e_i}^*(\theta_1,...,\theta_n) = \begin{cases} d_{e_i}(\theta_1,...,\theta_n) & e_i \neq \eta_i(\theta_i) \\ \\ f(\theta_i,\theta_{-i}) & e_i = \eta_i(\theta_i) \end{cases}$$

It is easy to see that since the evidence selection $(\eta_i)_{i\in N}$ satisfies condition (iii) in the definition of an incentive compatible evidence selection, so does $d_{e_i}^*$. Thus the maximally revealing evidence selection is incentive compatible.

Now consider Bayesian monotonicity. Observe that the normal evidence structure implies that whenever type θ_i can mimic type θ_i' under the maximally revealing evidence selection η_i^* (that is, whenever $\bar{e}_{\theta_i'} \in E_i^{\theta_i}$), he could also do so in the original evidence selection η_i. Thus condition (ii) in the definition of the Bayesian monotonic evidence structure $(\eta_i)_{i\in N}$ implies that a similar condition holds for evidence selection $(\eta_i^*)_{i\in N}$. We conclude that the maximally revealing evidence selection is also Bayesian monotonic.

4.4 Sufficient Condition

Proposition 26 provides conditions that a scf needs to satisfy in order to be fully implementable. In this section, instead, we will characterize sufficient conditions to attain full implementation. In particular, we will divide the analysis in two parts. In Section 4.4.1, we show that the EIC-EBM condition is also sufficient if we consider a specific, but very relevant class of environments; this result also provides a justification for the necessary condition provided in Proposition 26. In Section 4.4.2, instead, we consider general environments and we impose a further assumption (a no-veto assumption) that enable us to get the sufficiency result.[21] The gap between necessity and sufficiency in general environments is not specific

[21] In Jackson (1991)'s paper, a closure condition is also necessary for implementation. Since we are dealing with social choice functions, this condition is trivially satisfied. However, if we were to analyze the more general case of social choice rules, we should impose a similar closure condition.

to our own paper: most of the implementation literature has provided results in which a similar gap exists. The complete characterization of fully implementable scfs in general environments is an open question for future research.

4.4.1 Economic Environments

We start the analysis by defining what an economic environment is.

Definition 18 An environment $\mathcal{E}$ is economic if for every function $q : \Theta \to A$ and for every $\theta \in \Theta$, there exists a pair of agents i and j with $i \neq j$, and a pair of alternatives $x, y \in A$, such that for every set $C \subseteq \Theta_{-i}$ with $\theta_{-i} \in C_{-i}$:

$$\sum_{\theta_{-i} \in C_{-i}} u_i\left(x, \theta_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i\right) + \\ + \sum_{\theta_{-i} \in \Theta_{-i} \setminus C_{-i}} u_i\left(q\left(\theta_i, \theta_{-i}\right), \theta_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i\right) > \\ > \sum_{\theta_{-i} \in \Theta_{-i}} u_i\left(q\left(\theta_i, \theta_{-i}\right), \theta_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i\right)$$

and at the same time:

$$\sum_{\theta_{-j} \in C_{-j}} u_j\left(y, \theta_j, \theta_{-j}\right) \pi\left(\theta_{-j} \mid \theta_j\right) + \\ + \sum_{\theta_{-j} \in \Theta_{-j} \setminus C_{-j}} u_j\left(q\left(\theta_j, \theta_{-j}\right), \theta_j, \theta_{-j}\right) \pi\left(\theta_{-j} \mid \theta_j\right) > \\ > \sum_{\theta_{-j} \in \Theta_{-j}} u_j\left(q\left(\theta_j, \theta_{-j}\right), \theta_j, \theta_{-j}\right) \pi\left(\theta_{-j} \mid \theta_j\right)$$

Thus an environment is economic if for every state and for every mapping from states to alternatives the planner may choose, we can find at least two agents who would like to modify this rule for each possible subset of opponents' information. Intuitively, these environments are called economic because they involve a fundamental disagreement among agents.

The next Theorem shows that in an economic environment with at least 3 agents, the EIC-EBM condition fully characterizes the set of fully implementable social choice functions.

Theorem 28 Let $\mathcal{E}$ be an economic environment such that $|N| \geq 3$.[22] Assume that a social choice function f satisfies the EIC-EBM condition, then f is fully implementable.

Proof: Proposition 26 implies that the EIC-EBM is a necessary condition for implementation. Thus we only need to check the sufficiency part. To achieve this goal we will construct a mechanism that fully implements f. Consider a mechanism in which the message space of each agent is given by:

$$M_i = \Theta_i \times E_i \times \left(A^{\Theta} \cup \emptyset\right) \times A^{\Theta} \times \mathbb{N}$$

Thus a message for player i (m_i) will be denoted with the vector $(m_i^1, .., m_i^5)$, where m_i^j represents the j-th component of m_i (for example m_i^2 represents the evidence provided by agent i when he sends message m_i). Let $(\eta_i^* : \Theta_i \to E_i)$ be an evidence selection satisfying incentive compatibility and Bayesian monotonicity with respect to f. The existence of such a selection is guaranteed by the assumption that f satisfies the EIC-EBM condition. For every i and e_i, let d_{e_i} be the function that guarantees that η_i^* satisfies condition (iii) in the definition of an incentive compatible evidence selection with respect to f.

The outcome function of the mechanism, namely $g : M \to A$, is characterized considering 4 different cases.[23]

(1) if for every $i \in N$ $m_i = (.,.,\emptyset,.,.)$ and for every agent i, $m_i^2 = \eta_i^*(m_i^1)$, then $g(m) = f(m_1^1, ..., m_n^1)$;

(2) if for every $i \in N$ with $i \neq j$, $m_i(.,.,\emptyset,.,.)$ and $m_i^2 = \eta_i^*(m_i^1)$, while $m_j = (.,.,\emptyset,.,.)$ with $m_j^2 \neq \eta_j^*\left(m_j^1\right)$, then let $g(m) = d_{m_j^2}(m_1^1, ..., m_n^1)$;

(3) if for every $i \in N$ with $i \neq j$, $m_i(.,.,\emptyset,.,.)$ and $m_i^2 = \eta_i^*(m_i^1)$, while agent j sends

[22] Given a set X, $|X|$ denotes the cardinality of X.

[23] In building the mechanism, I will use the integer game that is standard in the implementation literature. We conjecture that the critique against this unbounded mechanism can be adressed using a bounded modulo game as in Jackson (1991). For a discussion of implementation with bounded mechanism, see Jackson (1992).

message $m_j = (.,.,y,.,.)$ then we can have 3 subcases:

(3(i)) if $m_j^2 \neq \eta_j^*\left(m_j^1\right)$, let $g(m) = d_{m_j^2}(m_1^1, ..., m_n^1)$;

(3(ii)) if $m_j^2 = \eta_j^*\left(m_j^1\right)$ and for some $\theta_j \in \Theta\left(m_j^2\right)$:

$$\sum_{\theta_{-j}} u_j\left(y\left(m_j^1, \theta_{-j}\right), \theta_j, \theta_{-j}\right)\pi\left(\theta_{-j} \mid \theta_j\right) > \sum_{\theta_{-j}} u_j\left(f\left(\theta_j, \theta_{-j}\right), \theta_j, \theta_{-j}\right)\pi\left(\theta_{-j} \mid \theta_j\right)$$

then $g(m) = f(m_1^1, m_2^1, ..., m_n^1)$;

(3(iii)) if $m_j^2 = \eta_j^*\left(m_j^1\right)$ and for every $\theta_j \in \Theta\left(m_j^2\right)$:

$$\sum_{\theta_{-j}} u_j\left(f\left(\theta_j, \theta_{-j}\right), \theta_j, \theta_{-j}\right)\pi\left(\theta_{-j} \mid \theta_j\right) \geq \sum_{\theta_{-j}} u_j\left(y\left(m_j^1, \theta_{-j}\right), \theta_j, \theta_{-j}\right)\pi\left(\theta_{-j} \mid \theta_j\right)$$

then implement $g(m) = y(m_1^1, m_2^1, ..., m_n^1)$;

(4) in any other case, define $i^* = \arg\max_i m_i^5$ and let $g(m) = m_{i^*}^4(m_1^1, m_2^1, ..., m_n^1)$.

We start showing that this mechanism has an equilibrium $\left(\sigma_i^* : \theta_i \to M_i\right)_{i \in N}$ such that for every θ, $g(\sigma^*(\theta)) = f(\theta)$. Consider the profile of strategies such that for every i and θ_i :

$$\sigma_i^*(\theta_i) = (\theta_i, \eta_i^*(\theta_i), \emptyset, ., .)$$

If this strategy were an equilibrium, the mechanism we described would result in function $f(.)$ being implemented (indeed we will show that EIC is all what we need to get this result). Pick an agent i and an information θ_i and consider all possible deviations:

(a) if he deviates to $(\theta_i, e_i, \emptyset, ., .)$, with $e_i \neq \eta_i^*(\theta_i)$, the outcome is determined by function d_{e_i} and condition (iii) in the definition of an incentive compatible evidence selection implies that this deviation is not profitable;

(b) if he announces $(\alpha_i(\theta_i), e_i, \emptyset, ., .)$, with $e_i = \eta_i^*(\alpha_i(\theta_i))$, condition (ii) in the definition of an incentive compatible evidence selection guarantees that this is not a profitable deviation;

(c) if he deviates to $(.,.,y,.,.)$, then we can have 3 subcases:

(c(i)) if $m_i^2 \neq \eta_i^*(m_i^1)$, the construction of function $d_{m_i^2}$ guarantees that this cannot be a

profitable deviation;

(c(ii)) if $m_i^2 = \eta_i^*(m_i^1)$ and for some $\theta_i \in \Theta(m_i^2)$:

$$\sum_{\theta_{-i}} u_i\left(y\left(m_i^1, \theta_{-i}\right), m_i^1, \theta_{-i}\right) \pi\left(\theta_{-i} \mid m_i^1\right) > \sum_{\theta_{-i}} u_i\left(f\left(m_i^1, \theta_{-i}\right), m_i^1, \theta_{-i}\right) \pi\left(\theta_{-i} \mid m_i^1\right)$$

$f(m_1^1, ..., m_n^1)$ is implemented and by condition (ii) in the definition of incentive compatible evidence selection, this is not a profitable deviation;

(c(iii)) if $m_i^2 = \eta_i^*(m_i^1)$ and for every $\theta_i \in \Theta(m_i^2)$

$$\sum_{\theta_{-i}} u_i\left(f\left(m_i^1, \theta_{-i}\right), m_i^1, \theta_{-i}\right) \pi\left(\theta_{-i} \mid m_i^1\right) \geq \sum_{\theta_{-i}} u_i\left(y\left(m_i^1, \theta_{-i}\right), m_i^1, \theta_{-i}\right) \pi\left(\theta_{-i} \mid m_i^1\right)$$

then $y(m_1^1, ..., m_n^1)$ and, by construction, this cannot be a profitable deviation.

Since the same reasoning holds for all agents and for all informations, we can conclude that this mechanism admits a Bayes-Nash equilibrium that implements f; therefore f can be partially implemented with this mechanism.

We will now show, that any other equilibrium of the mechanism induce outcome $f(\theta)$ in each state θ. Consider any equilibrium $(\sigma_i^*(.))_{i \in N}$ of the mechanism and let $\alpha = (\alpha_i(.))_{i \in N}$ be the profile of deceptions played by agents in this equilibrium.

We start observing that in this equilibrium it must be the case that for every i and θ_i, $\sigma_i^*(\theta_i) = (\alpha_i(\theta_i), \eta_i^*(\alpha_i(\theta_i)), \emptyset, ., .)$. Suppose this is not the case. Then either, $m_i^2 \neq \eta_i^*(\alpha_i(\theta_i))$ or $m_i^3 \neq \emptyset$. Let $\tilde{\Theta}$ be the set of states in which there is at least one agent s sending a message in which either $m_s^2 \neq \eta_s^*(\alpha_s(\theta_s))$ or $m_s^3 \neq \emptyset$ or both. Then, since the environment is economic, for every $\tilde{\theta} \in \tilde{\Theta}$ we can find two agents i and j and two alternatives x and y, such that i prefers x and j prefers y to whatever the mechanism is implementing in any subset of $\left\{\tilde{\theta}_j\right\} \times \tilde{\Theta}_{-j}$. Then when agent i has information $\tilde{\theta}_i$ (respectively, j has information $\tilde{\theta}_j$), he could announce $(., ., \emptyset, h_x, t)$ (respectively, $(., ., \emptyset, h_y, t)$) where t is a integer big enough and h_z is a constant function equal to z. By doing so, he could induce the implementation of

x in $\left\{\tilde{\theta}_i\right\} \times \tilde{\Theta}_{-i}$ (respectively y in $\left\{\tilde{\theta}_j\right\} \times \tilde{\Theta}_{-j}$) and increase his own utility. This deviation would not affect the outcome implemented outside $\tilde{\Theta}$. Thus, by construction, this represents a profitable deviation contradicting the assumption that σ^* is an equilibrium. We conclude that in any equilibrium of the mechanism we proposed, for every i and θ, $\sigma_i^*(\theta_i) = (.,.,\emptyset,.,.)$ with $m_i^2 = \eta_i^*(\alpha_i(\theta_i))$.

Now suppose that agents are playing a deception profile $\alpha = (\alpha_i)_{i \in N}$ such that $\alpha(\theta) \neq \theta$ for some θ and that for every i and for every information θ_i, $\sigma_i^*(\theta_i) = (\alpha_i(\theta_i), m_i^2, \emptyset, .,.)$ with $m_i^2 = \eta_i^*(\alpha_i(\theta_i))$. Since evidence selection $\eta^* = (\eta_i^*)$ is Bayesian Monotonic, we can find an agent i, an information θ_i^* and a function q such that:

$$\sum_{\theta_{-i}} u_i\left(q\left(\alpha_i\left(\theta_i^*\right), \alpha_{-i}\left(\theta_{-i}\right)\right), \theta_i^*, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i^*\right) > \\ > \sum_{\theta_{-i}} u_i\left(f\left(\alpha_i\left(\theta_i^*\right), \alpha_{-i}\left(\theta_{-i}\right)\right), \theta_i^*, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i^*\right)$$

while for every θ_i such that $\eta_i(\alpha_i(\theta_i^*)) \in E_i^{\theta_i}$:

$$\sum_{\theta_{-i}} u_i\left(f\left(\theta_i, \theta_{-i}\right), \theta_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i\right) \geq \sum_{\theta_{-i}} u_i\left(q\left(\alpha_i\left(\theta_i^*\right), \theta_{-i}\right), \theta_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i\right)$$

Then agent i in state θ_i^* can deviate to message $(\alpha_i(\theta_i^*), \eta_i^*(\alpha_i(\theta_i)), q, .,.)$ and induce outcome $q(\alpha_i(\theta_i^*), \alpha_{-i}(\theta_{-i}))$ to be implemented for any profile of information θ_{-i}. By the first inequality above this represents a profitable deviation contradicting the assumption that σ^* was an equilibrium. We conclude that the mechanism proposed fully implements $f(.)$.

The mechanism we construct in the proof of Proposition 28 allows the planner to use both the evidence structure and the conflict of interests among agents to kill bad equilibria. In particular, the outcome function of the mechanism is as follows. If all agents are sending messages in which (i) the evidence provided is compatible (according to the evidence selection chosen by the planner) with the information announced by the agent, and (ii) no agent claims that the others are playing a deception, the mechanism will implement the outcome

prescribed by the scf in the state announced by the agents. This is the case that arises in the equilibrium of the game. Off the equilibrium, several cases are possible. If a single agent provides an evidence, e_i, that does not match the information he claimed to have, the planner will punish him by using function d_{e_i}. On the other hand, if all agents send messages in which the evidence is compatible with the state announced, but a single agent raises a red flag claiming that the other agents are playing a deception, the planner will believe and reward him according to a function suggested by the agent himself, only if, *were the other agents being sincere,* he could never gain from having this reward function implemented (if this condition is not satisfied, the planner will ignore this tip). Finally if none of the previous cases applies, the mechanism prescribes an integer game that induces the agents to compete against each others and, thanks to the conflict of interests implied by the assumption of economic environments, prevents undesirable equilibria from arising.
In the special case in which there is no hard-evidence (for every i and every state θ, $E_i^\theta = E_i$), Theorem 28 replicates Theorem 1 in Jackson (1991).

4.4.2 General Environments

Although reasonable in many settings, the assumption of economic environments is not always valid. To address this issue, we now consider the problem of full implementation in a general class of environments in which the conflict of interests between agents may be less severe than what implied by an economic environment. This generality does not come for free: even if we are able to provide a sufficient condition to get full implementation, a gap between necessity and sufficiency arises. We start the analysis by introducing an assumption that will be used in the statement of the sufficient condition for full implementation.

Definition 19 A scf $f(.)$ satisfies the no-veto (NV) assumption for a feasible profile of deceptions α and a set $\hat{\Theta} \subseteq \Theta$, if for any $\theta \in \hat{\Theta}$, there exists i such that for every $j \neq i$ and

for every $q \in A^{\Theta}$, we can find a subset $\tilde{\Theta} \subseteq \hat{\Theta}$ such that $\theta \in \tilde{\Theta}$ and

$$\sum_{\theta_{-j} \in \Theta_{-j}} u_j \left(f\left(\theta_j, \theta_{-j}\right), \theta_j, \theta_{-j} \right) \pi\left(\theta_{-j} \mid \theta_j\right) \geq$$

$$\geq \sum_{\theta_{-j} \in \tilde{\Theta}_{-j}} u_j \left(q\left(\alpha_j\left(\theta_j\right), \alpha_{-j}\left(\theta_{-j}\right)\right), \theta_j, \theta_{-j} \right) \pi\left(\theta_{-j} \mid \theta_j\right) +$$

$$+ \sum_{\theta_{-j} \in \Theta_{-j} \setminus \tilde{\Theta}} u_j \left(f\left(\theta_j, \theta_{-j}\right), \theta_j, \theta_{-j} \right) \pi\left(\theta_{-j} \mid \theta_j\right)$$

A social choice function satisfies the NV assumption for a certain set $\hat{\Theta}$, if the previous condition holds for every profile of deceptions.

Intuitively, the no-veto assumption guarantees that the social choice function f is sufficiently robust against changes in information. It is satisfied for a set $\hat{\Theta}$ if, for for any possible deception α, we can find at least $n-1$ agents that prefers the outcome induced by the social choice function to the one induced by any alternative function as long as this function is used only on the subset $\hat{\Theta}$. This assumption is called no-veto because it implies that whenever a mapping between states and outcomes is the optimal one (in a certain subset) for $n-1$ agents, then function f must coincide with it; to put it differently, it states that the preferences of a single agent (against those of the remaining $n-1$ ones) cannot play an overwhelming role in determining the outcome prescribed by f.

Although the EIC-EBM and the NV assumptions play an important role to establish full implementation, imposing the two of them separately is not enough. Indeed, full implementation requires to combine the ideas behind incentive compatibility, Bayesian monotonicity and no-veto property in a particular way. We begin by combining Bayesian monotonicity and the no-veto assumption.

Definition 20 Given a scf f and an environment $\mathcal{E}$, we say that a profile of evidence selection $(\eta_i : \Theta_i \to E_i)_{i \in N}$ is no-veto Bayesian monotonic with respect to f if:

(1) for every i and θ_i, $\eta_i(\theta_i) \in E_i^{\theta_i}$;

(2) for every profile of deceptions α and set of states $B = (B_1, ..., B_n) \subseteq \times_{i \in N} \Theta_i$ for which we can find a function $g \in A^{\Theta}$ such that:

(i) for some $\hat{\theta} \in \Theta$, $g\left(\hat{\theta}\right) \neq f\left(\hat{\theta}\right)$;

(ii) if $\theta \in B$, $g(\theta) = f(\alpha(\theta))$;

(iii) $g(\theta)$ satisfies the no-veto hypothesis for deception α and set $\Theta \backslash B$;

then there must exists an agent i, an information $\theta_i^* \in B_i$ and a function $q \in A^{\Theta}$ such that:

$$\sum_{\theta_{-i} \in B_{-i}} u_i\left(q\left(\alpha_i\left(\theta_i^*\right), \alpha_{-i}\left(\theta_{-i}\right)\right), \left(\theta_i^*, \theta_{-i}\right)\right) \pi\left(\theta_{-i} \mid \theta_i^*\right) +$$

$$+ \sum_{\theta_{-i} \in (\Theta \backslash B)_{-i}} u_i\left(g\left(\theta_i^*, \theta_{-i}\right), \theta_i^*, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i^*\right) >$$

$$> \sum_{\theta_{-i}} u_i\left(g\left(\theta_i^*, \theta_{-i}\right), \left(\theta_i^*, \theta_{-i}\right)\right) \pi\left(\theta_{-i} \mid \theta_i^*\right)$$

while for every θ_i such that $\eta_i\left(\alpha_i\left(\theta_i^*\right)\right) \in E_i^{\theta_i}$:

$$\sum_{\theta_{-i}} u_i\left(f\left(\theta_i, \theta_{-i}\right), \theta_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i\right) \geq \sum_{\theta_{-i}} u_i\left(q\left(\alpha_i\left(\theta_i^*\right), \theta_{-i}\right), \theta_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i\right)$$

Although rather involved, the previous definition is closely related to the one of a Bayesian monotonic evidence selection and to the no-veto assumption. In particular, an evidence selection satisfies the NV-EBM condition if whistle-blowers are willing to send the tip when they are compensated only on a subset of possible states (the set B), while, everywhere else (namely, the set $\Theta \backslash B$) a function satisfying the no-veto assumption holds. This condition is necessary to deal with perverse situations in which agents may coordinate on a deception that, when revelaed by a whistle-blower, results in a reward function preferred by (at least) $n-1$ agents to the scf that the planner would like to implement. We can now combine all previous definitions and introduce a condition that will be sufficient to attain full implementation.

Definition 21 A scf f satisfies the EIC-NV-EBM condition if it admits a profile of evidence selection $(\eta_i)_{i\in N}$ that is both incentive compatible and no-veto Bayesian monotonic with respect to f.

The next Theorem states the role of the EIC-NV-EBM condition for full implementation in general environments.

Theorem 29 Consider an environment $\mathcal{E}$ with $|N| \geq 3$. If a scf f satisfies the EIC-NV-EBM condition, then it is implementable.

Proof: Consider the mechanism we used in the proof of Theorem 28. Following exactly the same steps of the proof in Theorem 28, we can show that the mechanism has an equilibrium that induces outcome $f(\theta)$ in every state θ. Therefore the mechanism we proposed partially implements social choice function f.

Let η^* be the evidence selection whose existence is asserted in the definition of EIC-NV-EBM. Suppose σ^* is an equilibrium of the mechanism such that for some θ, $g(\sigma^*(\theta)) \neq f(\theta)$. For every $(\theta_1, ..., \theta_n)$, let $(\sigma^*(\theta))^1 = \left((\sigma_1^*(\theta_1))^1, ..., (\sigma_n^*(\theta_n))^1\right)$ represents the information announced in this equilibrium. For every agent i, define the set:

$$B_i = \left\{\theta_i : \sigma_i^*(\theta_i) = \left((\sigma_i^*(\theta_i))^1, \eta_i^*\left((\sigma_i^*(\theta_i))^1\right), \emptyset, ., .\right)\right\}$$

Let $g(\sigma^*)$ represent the social choice function induced by the mechanism when agents follow strategy profile σ^* and define the deception profile $\alpha(.) = (\sigma^*(.))^1$. By the rules of the mechanism, for every $\theta \in B$, $g(\sigma^*(\theta)) = f(\alpha(.))$.

We start showing that $g(\sigma^*(.))$ must satisfy the no-veto assumption for the deception profile α and set $\Theta \setminus (\times_{i=1}^n B_i)$. Suppose this is not the case. Then, we can find a state

$\theta \in \Theta \backslash \left(\times_{i=1}^{n} B_i\right)$, two agents j', j'' and a function $q : \Theta \to A$ such that for every $j \in \{j', j''\}$:

$$\sum_{\theta_{-j} \in \tilde{\Theta}_{-j}} u_j \left(q\left(\alpha\left(\theta_j\right), \alpha_{-j}\left(\theta_{-j}\right)\right), \theta_j, \theta_{-j}\right) \pi\left(\theta_{-j} \mid \theta_j\right) + \\ + \sum_{\theta_{-j} \in \Theta_{-j} \backslash \tilde{\Theta}_{-j}} u_j \left(g\left(\sigma_j^*\left(\theta_j\right), \sigma_{-j}^*\left(\theta_{-j}\right)\right), \theta_j, \theta_{-j}\right) \pi\left(\theta_{-j} \mid \theta_j\right) > \\ > \sum_{\theta_{-j}} u_j \left(g\left(\sigma_j^*\left(\theta_j\right), \sigma_{-j}^*\left(\theta_{-j}\right)\right), \theta_j, \theta_{-j}\right) \pi\left(\theta_{-j} \mid \theta_j\right)$$

for all $\tilde{\Theta} \subseteq \Theta_{-j} \backslash \left(\times_{i=1}^{n} B_i\right)$ with $\theta \in \tilde{\Theta}$. Since $\theta \notin \left(\times_{i=1}^{n} B_i\right)$, there exists an agent s who is playing a message outside B_s. Then at least one of the agents $j \in \{j', j''\}$ could announce $(., ., \emptyset, q, t)$ when his information is θ_j. If t is sufficiently big, he can implement function q on $\Theta \backslash \left(\times_{i=1}^{n} B_i\right)$. By construction, this is a profitable deviation, contradicting the assumption that σ^* is an equilibrium. We conclude that $g\left(\sigma^*\left(.\right)\right)$ satisfies the no-veto assumption must be satisfied in $\Theta \backslash \left(\times_{i=1}^{n} B_i\right)$.

Now we want to show that for every θ, $g\left(\sigma\left(.\right)\right)$ has to be equivalent to $f\left(.\right)$. Suppose that this is not the case, then by the assumption EIC-NV-EBM (in particular by the assumption that the evidence selection is no-veto Bayesian monotonic), we can find an agent i, an information $\theta_i^* \in B_i$ and a state $q \in A^{\Theta}$ such that:

$$\sum_{\theta_{-i} \in B_{-i}} u_i \left(q\left(\alpha_i\left(\theta_i^*\right), \alpha_{-i}\left(\theta_{-i}\right)\right), \theta_i^*, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i^*\right) + \\ + \sum_{\theta_{-i} \in (\Theta \backslash B)_{-i}} u_i \left(g\left(\sigma_i^*\left(\theta_i^*\right), \sigma_{-i}^*\left(\theta_{-i}\right)\right), \theta_i^*, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i^*\right) > \\ > \sum_{\theta_{-i}} u_i \left(g\left(\sigma_i^*\left(\theta_i^*\right), \sigma_{-i}^*\left(\theta_{-i}\right)\right), \theta_i^*, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i^*\right)$$

and for every θ_i such that $\eta_i\left(\alpha_i\left(\theta_i^*\right)\right) \in E_i^{\theta_i}$:

$$\sum_{\theta_{-i}} u_i \left(f\left(\theta_i, \theta_{-i}\right), \theta_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i\right) \geq \sum_{\theta_{-i}} u_i \left(q\left(\alpha_i\left(\theta_i^*\right), \theta_{-i}\right), \theta_i, \theta_{-i}\right) \pi\left(\theta_{-i} \mid \theta_i\right)$$

Given these inequalities, when agent i has information θ_i^*, he can deviate and announce $(.,.,q,g(\sigma^*),t)$, where t is a sufficiently large integer. By doing so, he would be able to implement q in $\{\theta_i\} \times B_{-i}$ and $g(\sigma^*)$ in $\{\theta_i\} \times (\Theta_{-i} \backslash B_{-i})$. By construction this would represent a profitable deviation. So we reached a contradiction with the assumption that σ^* is an equilibrium. Thus the outcome induced by σ^* is equal to $f(\theta)$ for every state θ. Since this holds for any equilibrium we conclude that the proposed mechanism fully implements f.

Theorem 29 extends Theorem 2 in Jackson (1991) to environments in which evidence is available; in the special case of no hard evidence, our result coincides with Jackson (1991)'s one. Furthermore, also in this case, if the evidence structure is normal, the only evidence selection we need to check is the one entailing maximal evidence provision.[24]

Once more, we want to stress that Theorem 29 leaves a gap between the necessary and sufficient conditions required to fully implement a social choice function in general environments. Intuitively, this gap arises because, in general environments, we are not making any assumption about agents' disagreement. Indeed, EBM provides conditions under which deceptions can be identified either through the evidence provided or through the preferences of some agents. However, in the mechanism we are proposing, if an agent i is sending an out-of-equilibrium message (that is, a message in which either the evidence provided is not the one required by the planner to certify the information announced, or a deception is announced by sending a message m_i in which $m_i^3 \neq \emptyset$), then the outcome function induced by the mechanism could still be optimal for agent i.[25] If also the other agents find the choice rule implemented by the mechanism optimal, the planner may have a hard time in destroying this equilibrium. The existence of a no-veto Bayesian monotonic evidence selection takes

[24] Since the definition of the no-veto assumption is not affected by the evidence structure, the proof of this result is almost identical to the one of Proposition (27) and it is omitted.

[25] This could happen if other agents were also playing according to some deception. Observe that Condition (iii) in the definition of an incentive compatible evidence selection rules out this case, only if the other agents are being thruthful about their information.

care of this situation by assuming that he can indeed find an agent i and a reward function q such that for some information agent i will indeed be willing to act as a whistle-blower.
In economic environments, this problem is avoided because, the planner can always find an agent who is willing to (and, given the specific rules of the mechanism, is also able to) destroy this equilibrium. Indeed, when an environment is economic, the no-veto condition can never be satisfied and so EBM and EBM-NV coincide.
The previous discussion also sheds some light on the condition that would be required to guarantee a full characterization of the implementation problem in general environments: such a condition would probably involve a strengthening of EBM in all those states of nature in which the conflict of interest among agents and the evidence structure are not powerful enough to reveal the coordination of agents on a bad equilibrium.

4.5 Conclusion

In this paper, we looked at the problem of full Bayesian implementation with hard evidence, that is full implementation in Bayes-Nash equilibrium in situations where agents can bring non-counterfeitable evidence to support their cheap talk claims. Although in the paper we focused on the implementation of social choice functions, we conjecture that the extension to social choice correspondences would not entail any conceptual difficulty.[26]
In particular, we have provided a condition, called EIC-EBM, that is necessary and sufficient for full implementation in economic environments with at least 3 agents. This conditions can be regarded as the natural extension of incentive compatibility and Bayesian monotonicity to environments in which evidence provision is available. For general environments, the EIC-EBM condition is still necessary, but it is no longer sufficient. Nevertheless we characterized a condition, EIC-NV-EBM, that is sufficient to attain full Bayesian implementation. Providing

[26] Arguably, a mechanism closely related to the one constructed in this paper should attain this goal. The biggest difference would come from the enlargement of the message spaces so that agents need to specify a particular selection from the social choice correspondence that the planner wants to implement.

a tighter characterization of full Bayesian implementation in general environments represents an interesting direction for future research.[27]

In this paper, we dealt with implementation in pure strategy on environments with more than 3 agents. Relaxing either of these assumptions represents another possible topic that deserve further analysis.

The way in which we modeled hard evidence can be seen as the limit case in which the cost of forgery goes to infinity. The analysis of the interaction between the cost of forgery and the implementation problem is another interesting direction for future research. Kartik and Tercieux (2011) analyzed the complete information version of this problem and we believe that their insight can be adapted to the incomplete information case.

We conclude the paper with a short comment on the relevance of the results contained in this paper. Admittedly, these results are based on rather complicated definitions and we conjecture that a full characterization of the implementation problem in general environments would require even more involved definitions. This is somehow disappointing: whereas Maskin (1999) provides intuitive conditions that are easy to understand and to check, the problem of implementation in more general settings requires conditions that are hard to evaluate and to understand. Although we are sympathetic with the idea behind these critiques, we want to claim that this problem is not specific to our analysis. Jackson (1991)'s IC, BM and BM-NV are also rather involved and their meaning can be fully understood only in relation with the mechanism used to attain full implementation. More in general, we believe that the implementation problem studied in this paper should be seen as a first step: it establishes boundaries to the set of social choice functions that a planner can hope to accomplish. There is no reason why these boundaries should be easy to state, but they may still play an important role in leading the modeler to make additional (and more intuitive)

[27]The implementation literature is often characterized by this gap between necessary and sufficient conditions. One of the reasons for the existence of these gaps is that the definitions required for full characterization are often rather involved and far from intuitive. In the specific environment that we are considering we acknowledge that conditions like EIC-EBM, NV-EBM or EIC-NV-EBM are also rather involeved.

restrictions. To put it differently, we should first establish what can be possibly achieved and only then worry about what can be plausibly and more easily obtained.

Bibliography

Abreu, D., and H. Matsushima (1991): "Virtual Implementation in iteratively undominated strategies: incomplete information," *Mimeo*.

——— (1992a): "A response to Glazer and Rosenthal," *Econometrica*, 60(6), 1439–1442.

——— (1992b): "Virtual implementation in iteratively undominated strategies: complete information," *Econometrica*, 60(5), 993–1008.

Abreu, D., and A. Sen (1991): "Virtual implementation in Nash equilibrium," *Econometrica*, 59(4), 997–1021.

Aumann, R. J., and S. Hart (2002): "Long Cheap Talk," *Econometrica*, 71, 1619–1660.

Banks, J. S. (1990): "A model of electoral competition with incomplete information," *Journal of Economic Theory*, 50(2), 309–325.

Banks, J. S., and R. K. Sundaram (1998): "Optimal Retention in Agency Problems," *Journal of Economic Theory*, 82(2), 293–323.

Battaglini, M. (2002): "Multiple Referrals and Multidimensional Cheap Talk," *Econometrica*, 70(4), 1379–1401.

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

——— (2009): "Dynamic psychological games," *Journal of Economic Theory*, 144(1), 1–35.

Ben-Porath, E., and B. Lipman (2011): "Implementation and Partial Provability," *Journal of Economic Theory (Forthcoming)*.

Benoit, J., and E. Ok (2006): "Maskin's Theorem with limited veto power," *Games and Economic Behavior*, 55(2), 331–339.

Berganza, J. C. (2000): "Two Roles for Elections: Disciplining the Incumbent and Selecting a Competent Candidate," *Public Choice*, 105, 165–194, 10.1023/A:1005258810057.

Bergemann, D., and S. Morris (2005): "Robust mechanism design," *Econometrica*, 73(6), 1771–1813.

——— (2008a): "Ex post implementation," *Games and Economic Behavior*, 63(2), 527–566.

——— (2008b): "The role of the common prior in robust implementation," *Journal of the European Economic Association*, 6(2-3), 551–559.

——— (2009a): "Robust implementation in direct mechanisms," *Review of Economic Studies*, 76(4), 1175–1204.

——— (2009b): "Robust virtual implementation," *Theoretical Economics*, 4(1), 45–88.

——— (2010): "Robust implementation in general mechanisms," *Games and Economic Behavior*, 71(2), 261–281.

Bergemann, D., S. Morris, and O. Tercieux (2011): "Rationalizable implementation," *Journal of Economic Theory*, 146(3), 1253–1274.

Brunnermeier, M. K., and J. A. Parker (2005): "Optimal Expectations," *American Economic Review*, 95(4), 1092–1118.

Bull, J., and J. Watson (2004): "Evidence disclosure and verifiability," *Journal of Economic Theory*, 118(1), 1–31.

——— (2007): "Hard evidence and mechanism design," *Games and Economic Behavior*, 58(1), 75–93.

Caplin, A., and J. Leahy (2001): "Psychological Expected Utility Theory And Anticipatory Feelings," *The Quarterly Journal of Economics*, 116(1), 55–79.

Charness, G., and M. Dufwenberg (2006): "Promises and Partnership," *Econometrica*, 74(6), 1579–1601.

——— (2011): "Participation," *American Economic Review*, 101(4), 1211–37.

Crawford, V. P., and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50(6), 1431–51.

de Meza, D., and D. C. Webb (2007): "Incentive Design Under Loss Aversion," *Journal of the European Economic Association*, 5(1), 66–92.

Downs, A. (1957): "An Economic Theory of Political Action in a Democracy," *Journal of Political Economy*, 65(2), pp. 135–150.

Duggan, J. (2000): "Repeated Elections with Asymmetric Information," *Economics and Politics*, 12(2), 109–135.

Duggan, J., and M. Fey (2006): "Repeated Downsian electoral competition," *International Journal of Game Theory*, 35, 39–69, 10.1007/s00182-006-0046-1.

Dziuda, W. (2011): "Communication with Detectable Deceit," Mimeo.

Epstein, L. G. (2008): "Living with Risk," *Review of Economic Studies*, 75(4), 1121–1141.

Farejohn, J. (1986): "Incvumbent Performance and Electoral Control," *Public Choice*, 50(1-3), 5–25.

Farrell, J. (1995): "Talk is Cheap," *American Economic Review*, 85, 186–190.

Farrell, J., and R. Gibbons (1989): "Cheap Talk with Two Audiences," *American Economic Review*, 79(5), 1214–23.

Farrell, J., and M. Rabin (1996): "Cheap Talk," *Journal of Economic Perspectives*, 10(3), 103–18.

Fehr, E., O. Hart, and C. Zehnder (2011): "Contracts as Reference Points–Experimental Evidence," *American Economic Review*, 101(2), 493–525.

Geanakoplos, J., D. Pearce, and E. Stacchetti (1989): "Psychological games and sequential rationality," *Games and Economic Behavior*, 1(1), 60 – 79.

Glazer, J., and R. Rosenthal (1992): "A note on Abreu-Matsushima mechanisms," *Econometrica*, 60(6), 1435–1438.

Goltsman, M., J. Horner, G. Pavlov, and F. Squintani (2009): "Mediation, arbitration and negotiation," *Journal of Economic Theory*, 144(4), 1397–1420.

G.R. Boynton, R. D. H., and S. C. Patterson (1969): "Perceptions and Expectations of the Legislature and Support for It," *American Journal of Sociology*, 75(1), 62–76.

Green, J. R., and N. L. Stokey (2007): "A two-person game of information transmission," *Journal of Economic Theory*, 135(1), 90–104.

Grillo, E. (2011a): "Electoral Speeches and Reference Dependence: Can Politicians' Announcement be Credible?," Mimeo.

——— (2011b): "Reference-Dependence, Risky Projects and Credible Information Transmission," Mimeo.

Gul, F. (1991): "A Theory of Disappointment Aversion," *Econometrica*, 59(3), pp. 667–686.

Hart, O., and J. Moore (2008): "Contracts as Reference Points," *The Quarterly Journal of Economics*, 123, 1–48.

Hotelling, H. (1929): "Stability in Competition," *The Economic Journal*, 39(153), pp. 41–57.

Ivanov, M. (2010): "Communication via a strategic mediator," *Journal of Economic Theory*, 145(2), 869–884.

Jackson, M. (1991): "Bayesian implementation," *Econometrica*, 59(2), 461–477.

——— (1992): "Implementation in undominated strategies: A look at bounded mechanisms," *The Review of Economic Studies*, 59(4), 757–775.

James, O. (2009): "Evaluating the Expectations Disconfirmation and Expectations Anchoring Approaches to Citizen Satisfaction with Local Public Services," *Journal of Public Administration Research and Theory*, 19(1), 107–123.

Kahneman, D. (1992): "Reference points, anchors, norms, and mixed feelings," *Organizational Behavior and Human Decision Processes*, 51(2), 296–312.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1990): "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy*, 98(6), 1325–48.

——— (1991): "The Endowment Effect, Loss Aversion, and Status Quo Bias: Anomalies," *Journal of Economic Perspectives*, 5(1), 193–206.

Kahneman, D., and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47(2), 263–91.

Kartik, N. (2009): "Strategic Communication with Lying Costs," *Review of Economic Studies*, 76(4), 1359–1395.

Kartik, N., M. Ottaviani, and F. Squintani (2007): "Credulity, lies, and costly talk," *Journal of Economic Theory*, 134(1), 93–116.

Kartik, N., and O. Tercieux (2011): "Implementation with evidence," *Theoretical Economics (Forthcoming)*.

Kimball, D. C., and S. C. Patterson (1997): "Living up to Expectations: Public Attitudes toward Congress," *Journal of Politics*, 59(3), 701–728.

Koszegi, B. (2006): "Emotional Agency," *The Quarterly Journal of Economics*, 121(1), 121–156.

——— (2010): "Utility from Anticipation and Personal Equilibrium," *Economic Theory*, 44(3), 415–444.

Koszegi, B., and M. Rabin (2006): "A Model of Reference-Dependent Preferences," *The Quarterly Journal of Economics*, 121, 1133–1165.

——— (2007): "Reference-Dependent Risk Attitudes," *American Economic Review*, 97(4), 1047–1073.

——— (2009): "Reference-Dependent Consumption Plans," *American Economic Review*, 99(3), 909–36.

Krishna, V., and J. Morgan (2001): "A Model Of Expertise," *The Quarterly Journal of Economics*, 116(2), 747–775.

——— (2004): "The art of conversation: eliciting information from experts through multi-stage communication," *Journal of Economic Theory*, 117(2), 147–179.

——— (2008): "Cheap Talk," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume. Palgrave Macmillan, Basingstoke.

Lindstadt, R., and J. K. Staton (2010): "Managing Expectations: When Does it Work and What are the Consequences?," Mimeo.

Loewenstein, G. (1987): "Anticipation and the Valuation of Delayed Consumption," *Economic Journal*, 97(387), 666–84.

Loewenstein, G., and D. Prelec (1992): "Anomalies in Intertemporal Choice: Evidence and an Interpretation," *The Quarterly Journal of Economics*, 107(2), pp. 573–597.

Maskin, E. (1999): "Nash Equilibrium and Welfare Optimality*," *The Review of Economic Studies*, 66(1), 23–38.

Moore, J., and R. Repullo (1988): "Subgame perfect implementation," *Econometrica*, 56(5), 1191–1220.

——— (1990): "Nash implementation: a full characterization," *Econometrica*, 58(5), 1083–1099.

Myerson, R. (1982): "Optimal coordination mechanisms in generalized principal-agent problems," *Journal of mathematical economics*, 10(1), 67–81.

Ottaviani, M., and F. Squintani (2006): "Naive audience and communication bias," *International Journal of Game Theory*, 35(1), 129–150.

Oury, M., and O. Tercieux (2011): "Continuous Implementation," *Econometrica (Forthcoming)*.

Palfrey, T., S. Srivastava, and A. Postlewaite (1993): *Bayesian implementation*, vol. 1. CRC.

Penta, A. (2011): "Robust Dynamic Mechanism Design," Mimeo.

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

Richard W. Waterman, H. C. J.-S., and C. L. Silva (1999): "The Expectations Gap Thesis: Public Attitudes toward an Incumbent President," *Journal of Politics*, 61(4), 944–966.

Schwabe, R. (2011): "Reputation and Accountability in Repeated Elections," Mimeo.

Shalev, J. (2000): "Loss aversion equilibrium," *International Journal of Game Theory*, 29(2), 269–287.

Sigelman, L., and K. Knight (1983): "Why Does Presidential Popularity Decline? A Test of the Expectation/Disillusion Theory," *Public Opinion Quarterly*, 47(3), 310–324.

Sugden, R. (2003a): "Reference-dependent subjective expected utility," *Journal of Economic Theory*, 111(2), 172–191.

——— (2003b): "Reference-dependent subjective expected utility," *Journal of Economic Theory*, 111(2), 172–191.

Tversky, A., and D. Kahneman (1991): "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *The Quarterly Journal of Economics*, 106(4), 1039–61.

van Dijk, E., and D. van Knippenberg (1996): "Buying and selling exchange goods: Loss aversion and the endowment effect," *Journal of Economic Psychology*, 17(4), 517 – 524.