ESSAYS ON IMPERFECT INFORMATION PROCESSING IN ECONOMICS

ISBN 978 90 5170 775 5

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. **406** of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Essays on Imperfect Information Processing in Economics

Essays over imperfecte informatie verwerking in de economie

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. S.W.J. Lamberts

and in accordance with the decision of the Doctorate Board

The public defence shall be held on
Friday 25 May 2007 at 11.00 hrs

By

Stefano Salvatore Ficco

born in Milan, Italy

Doctoral Committee

Promoter:	Prof.dr. M. C. W. Janssen
Other members:	Prof.dr. H. Bleichrodt Prof.dr. O. H. Swank Prof.dr. T. Van Zandt

Preface

This thesis is the result of the doctoral research I conducted at the Tinbergen Institute and at the Department of Economics of the Erasmus University in Rotterdam. Several people have contributed to the successful completion of this dissertation.

First and foremost, I wish to thank my promoter Maarten Janssen. I first met Maarten in 1998 when I was in Rotterdam as an exchange student, and it is thank to him that I had the opportunity to come back three years later to start my Ph.D. track. In these years Maarten has been an outstanding supervisor, he provided me with guidance and support (both material and psychological), and always helped me to overcome the difficulties encountered along the way.

I also wish to express my deepest gratitude to Vladimir Karamychev. Without his patience and priceless help this achievement would have never been possible. Besides being the best coauthor one could ever have, Vladimir has been to me much more than that.

I also thank Han Bleichrodt, Otto Swank and Timothy Van Zandt, for being part of the Doctoral Committee, and Robert Dur and Bauke Visser for their role in the defence. I am grateful to all the staff of the Tinbergen Institute and the Department of Economics, and in particular to Jose' Luis Moraga and Peter Wakker for their valuable comments.

I would also like to spend a few words of gratitude to anyone who made my stay in Rotterdam somehow special. Thank to Gerrit, with whom I shared the office for four years, for being a true and faithful friend. Thank to Chen, Francesco and Haris for helping me in my teaching duties. My affection to Alessandro, Andrea, Bobo, Caterina, Francesco, Jaap, Leon, Maria, Marielle, Rob and Sandra for giving me the feeling of having always someone to rely on.

Finally, I want to express my gratitude to Isjah for sharing with me this special moment, and to my brother Marco, my parents, and my grandmother, for the time and affection they deserve and that I was not able to give them in the last years.

Table of contents

1 Imperfect information processing in economics

2 Markets with information overload

3 Collective decision making and imperfect information processing

1. Imperfect information processing in economics

1 Introduction

Economic agents generally operate in uncertain environments and, prior to making decisions, invest time and resources to collect useful information. Consumers compare the prices charged by different firms before purchasing a product. Politicians gather information from different sources before implementing a policy. Doctors conduct several tests on patients before subscribing a treatment. Since having access to valuable information requires time, effort, and often money, a problem typically addressed by economists is how much information a decision-maker should collect before being sufficiently convinced about one state of the world. Implicit in economic models that aim to identify the optimal amount of information to be gathered is the idea that, once information is available, there are no costs involved in interpreting and transmitting it. However there are plausible reasons to believe that the activity of information processing is far from being perfect.

Looking at the quantitative dimension of information, a larger amount of information usually implies that more resources are needed in order to interpret it. In this respect, with the rise of the Internet, the cost of generating and transmitting information has fallen so dramatically that the scarce resource is not anymore information itself, but the attention needed to understand it. For example, the cost of sending an electronic manuscript to a journal is much lower than the opportunity cost of the time it takes to a referee to read and understand the paper. Modelling explicitly such opportunity costs may shed light on phenomena such as market congestion and information overload.

From a qualitative point of view, information may have a high degree of complexity which makes it not equally accessible to all economic agents. For instance, specialization of labor is nowadays so high that communication problems between agents with different competencies are very likely to arise. Suppose a car manufacturer has to choose one from many potential prototype projects in order to launch a new car model. If the selection criteria must pay attention to both technical (performance) and esthetical (design) considerations, it becomes crucial for the quality of the decision, whether a report compiled by the Engineering Department can be effectively understood by the Marketing&Sales Department, and vice versa. It seems then important to analyze models that relate the performance of collective decision making to the communication flow within the organization.

Finally, the idea that the human brain has a limited information processing capability is widely accepted in psychology. In this respect, the need for economic models in which the decision-makers' cognitive limitations are explicitly taken into account was already pointed out in the pioneering work of Simon (1955) on bounded rationality. However only recently the notion of bounded rationality has influenced mainstream economics and the state of the art in this field of research is far from being established. Consequently, more theoretical contributions to this branch of the literature are needed.

In what follows I will survey some recent works in which imperfect information processing is related to, respectively, the functioning of markets, collective decision making, and individual decision-making. I will discuss the main findings for each subject considered, and conclude by introducing the contribution that the present thesis provides in each of the aforementioned issues.

2 Markets and imperfect information processing

Within each market interaction a certain amount of information flows from some agents to others. Firms attract the attention of consumers by targeting them with advertisements that include information about their products, news about price discounts and special offers. Economists compete for publications by submitting to journals manuscripts that need to be processed by referees. In the job market the information included in application letters is sent from potential employees to the employer. In all these situations imperfect information processing typically refers to the inability of the agents in charge of screening information to effectively process all information received. This is a fairly realistic assumption as the cost of transmitting information has decreased so drastically, that the relatively scarce resource has become the human attention needed to process information.

In psychology, the idea that there exists limits on the information processing capability of the human brain is a well established fact. In his seminal paper, Miller (1965) pointed out that it is the span of absolute judgment and the span of immediate memory that impose severe limitations on the amount of information we are able to simultaneously process and remember. Libowsky (1975) is concerned with the psychophysiological consequences of experiencing overload on a daily basis, and argues that this phenomenon is a collateral consequence of both man-made transformation of the physical environment, and the mass production of symbols and messages facilitated by the growth of information technology. Yet, in economics very little attention has been paid to formalize the concept of information overload. One reason for this is the traditional argument by which information is regarded as a freely disposable good: excessive information can always be disregarded, therefore the mere fact that someone receives more information than he can process does not mean that he receives too much information. However there are empirical findings suggesting the opposite. In the experimental marketing literature, Jacoby et al. (1974) find that the accuracy which consumers make a right purchase is a bell-shaped function of the amount of information at disposal: too little as well as too much information results in a poor decision, implying that, in principle, there exists an optimal intermediate level of information to process. Surprisingly their experiments show that, despite consumers that were provided with more information made poorer decisions, they also felt more satisfied and less confused. Such finding suggests that consumers may not able to shield themselves from being overloaded when too much information is made readily available. Similar empirical evidence is provided by Malhotra (1982), and Keller and Staelin (1987).

Even if we retain the idea that individuals are able to accurately perceive their objective information processing limits, inefficiencies may still arise as a consequence of the fact that human attention is an unpriced resource. For example, a recent laboratory experiment by Kraut et. al. (2002) finds that charging for an e-mail may improve the effectiveness of communication.

To my knowledge, the only theoretical treatment of information overload in economics is Van Zandt (2004). The aforementioned paper addresses information overload issues for a network of targeted communication: several firms (senders) send advertisement messages to an arbitrary subset of consumers (receivers). There exists an informational asymmetry between senders, who know the content of their messages but are not fully aware about the interests of potential receivers, and receivers, who know their own interests but ignore the content of a message unless they process it. Limited attention is modelled by assuming that receivers can process only a fixed amount of messages. If they receive more messages than they can process, they randomly select a number of them without knowing their content. The main result of the paper is that information overload is a consequence of the strategic interaction among senders. Since the attention of receivers is an unpriced resource, each sender's message crowds out messages of other senders, and an increase in the amount of information results in a decrease in the average value of the information. Since the focus in Van Zandt (2004) is on targeted communication networks, the leading interpretation of the model is as advertising and a theoretical treatment of information overload in different contexts is left to future research.

A market structure that seems to me particularly relevant to address information overload issues is that of monopsony market. Consider for instance the following example: the receiver is a university posting one job opening on the Internet with exogenously fixed terms, and senders are applicants that decide whether to send their CV's. The limited time span in which such hiring decisions are typically taken, as well as the limited budget that universities have to evaluate all available job applicants, suggest that information overload problems in these contexts may arise. This motivates chapter 2 of the present thesis.

In chapter 2 I develop a model in which heterogenous senders (job applicants) compete in order to be selected by one receiver (employer). Productivity is private information to the senders, and the receiver processes imprecise informative signals about workers' productivity (applications) in order to screen among applicants. Limited attention is modeled by assuming that the information processing technology is imperfect: the accuracy of each signal in predicting the unknown productivity decreases with the total number of applications processed. With this setup the crucial decision variable for the receiver is to decide the actual number of applications to process.

My first finding is that, for any information processing technology level, information overload emerges when the cost of sending an application is low relatively to the existing technology level. Information overload is defined here as an equilibrium in which some of the applications sent are disregarded by the receiver. The receiver's ability to process an optimal amount of information may suggest the idea that receiving information that exceeds the processing capacity does not mean receiving too much information. However, I find that a lower sending cost is always associated to a larger amount of information disregarded in equilibrium and a lower average quality of the information. This result is supported by the empirical findings of Kraut et al. (2002) and confirms the analytical results of Van Zandt (2004).

The second issue I address in Chapter 2 is the effect of market size on information overload. Surprisingly, I find that, if the size of the market (that is, the number of senders) increases, the amount of information neglected in equilibrium stays constant and the average quality of information increases. This result is driven by the self-selection mechanism that typically operates in screening markets: the increase in competition discourages low-productive senders from applying and, therefore, a larger market translates only in higher average productivity of applicants.

Finally, I investigate the effect of relaxing the assumption that information is freely disposable, as it allows me to capture situations in which neglecting information is not possible. Such impossibility may be due to legal constraints or to specific policy orientations (for instance is common for universities to give all candidates the opportunity to present a job market paper). With this setup, information overload refers to the difference between the total number of applications sent and the number of applications that would be optimal to process. I find that the inefficiencies of information overload are more severe in a market where information is not freely disposable. Besides the fact that excessive information now directly harms the receiver who (contrary to the case in which information is freely disposable) cannot shield himself form being overloaded, I am also able to show that the inability of disregarding information results in a larger amount of information being sent as well as a lower average quality of such information.

3 Collective decision-making and imperfect information processing

The conventional economic paradigm views economic organizations as single decision-making entities. For instance, the behavior of a profit maximizing firm is typically modeled by assuming that a manager chooses the output level optimally. Yet, the decision-making process of a firm is in reality very complex and it involves many people. Consequently, in the last couple of decades, the research agenda has identified the need of modelling explicitly collective decision making as a necessary step in order to capture insights of how organizations work. The relevance of economic theories of organizations is implicitly justified by the fact that the activity of information processing is imperfect.

First, the limited capabilities of individuals to process information within a limited amount of time suggests that groups of individuals may be able to make better decisions than any single individual. Second, if individuals make systematic errors in interpreting the information they posses, the specific organizational structure in which they are embedded may affect the outcome of the final decision. Since the basic feature of an organization is that of aggregating decisions taken by each one of its members separately, different organizational structures will generally aggregate individual errors differently. Finally, the information handled by firms is a complex multidimensional entity. For instance, a decision about the investment opportunities in a specific country requires considerations of very different nature like, political stability, general economic figures, national fiscal regime, industry concentration etc. This is one reason why organizations decentralize the information processing activity by allocating the task of processing qualitatively different pieces of information to people with different competencies. The specialization of labor is nowadays so high that communication problems between agents with different backgrounds are very likely to arise. Consequently, the performance of collective decision making may be reduced due to imperfect communication flow within an organization.

Collective decision making is usually modelled by assuming that the organization is an information processing network in which each node represents an agent belonging to the organization. The majority of authors take the work of Marschak and Radner (1972) as a point of departure: agents are team members, that is, they have identical preferences and therefore agency problems do not arise[1]. Broadly speaking, contributions to this literature can be divided into two branches.

The first branch includes models in which the organization is neutral in the sense that the decision eventually reached by the organization is not

[1]In contrast, Lambert (1986) analyzes incentives problems in organization by taking a principal-agent approach. Incentives problems are partially addressed also in Koh (1992b).

affected by its internal structure. Neutrality arises whenever it is assumed that communication and aggregation do not affect the informational value that can be extracted from a batch of information. In all these models imperfect information processing refers to the fact that each agent has limited attention and, therefore, can process only a limited amount of information items within a time unit. Bolton and Dewatripont (1994) are concerned with the design of an efficient organizational structure when there is a trade-off between returns to specialization and cost of communication. On one hand, more specialization implies that agents have a lower unit time of processing a specific type of information. On the other hand, the more specialized agents are, the more communication is necessary to coordinate agents' activities. Radner (1992, 1993) initiated the research of decentralized information processing where organizations are viewed as a hierarchical structures that perform parallel associative computations. A similar approach is taken by Meagher and Van Zandt (1998), Radner and Van Zandt (1992, 2001), and Van Zandt (1998, 1999). Within this framework, the predominant focus of the analysis is to determine the efficient information processing system, that is, the one which performs the given task in the shortest possible time and by using the smallest possible number of agents.

A second branch of the literature (to which chapter 3 of the present thesis also belongs) includes models where, in contrast, the specific organizational structure affects the outcome of the decision-making process. The seminal work in this area is due to Sah and Stiglitz (1985, 1986). Their standpoint is human fallibility: individuals make systematic (honest) errors when making decisions and, therefore, the performance of a specific organization depends on the way in which it aggregates the errors made by each one of its agents separately. The focus is on the organization's decision about accepting or rejecting a certain project. Projects of ex-ante unknown quality are assumed to come in streams, and the organization evaluates them sequentially in order to distinguish those that are worth implementing from those that are not. The analysis is restricted to the comparison of two stylized organizational structures, a hierarchy and a polyarchy. In a hierarchy a project is accepted if and only if it has the approval of all its members, while for a project to be accepted in a polyarchy, the approval of a single member is sufficient. Since a hierarchy minimizes the errors of rejecting good projects (Type-I errors), while a polyarchy minimizes the errors of accepting bad projects (Type-II errors), the former organization performs better in a hostile environment[2], while the latter is more effective in a friendly environment. The human fallibility approach introduced by Sah and Stiglitz has inspired numerous authors that addressed other important organizational issues.

One such an issue is whether or not there exist organizational structures

[2] An environment is hostile (friendly) if the expected payoff from accepting a project at random is negative (positive).

that are more efficient than a pure hierarchy or a pure polyarchy. Ioannides (1987) shows that the screening performance can be increased if the organization adopts a complex hybrid form composed by both hierarchies and polyarchies. In contrast, Koh (1992a) notices that if, in addition to the fixed cost of hiring screeners, one also takes into account the variable costs of evaluation, then a pure hierarchy or a pure polyarchy may be more efficient than hybrid forms. Moreover, Koh (1993) suggests that, where there exists a first-mover advantage, small and decentralized organizations like polyarchies may be preferred as the advantage of a quicker decision offsets the loss in screening accuracy.

Other authors have studied the relation between organizational forms and market structure. Bull and Ordover (1987) find that the optimal size of a polyarchy depends on the degree of competition in the product market and, moreover, that a competitive market leads to long-run efficiencies that do not arise in more concentrated markets. Hendrikse (1992) shows that both hierarchies and polyarchies may emerge in equilibrium, and that more competition is associated with more centralized hierarchical structures.

Human fallibility, as it is modeled by Sah and Stiglitz, does not leave room for individual rationality as agents are characterized by an exogenous pair of probabilities (screening functions) with which they accept good and bad projects. In this respect, Koh (1992b) generalizes the setup of Sah and Stiglitz by modelling agents as homogenous rational screeners who cannot fully communicate the information they have. Each agent observes an imprecise informative signal about a projects' quality, and chooses the optimal cutoff point, say r^*, such that he recommends the project if the corresponding signal is greater than r^* and rejects it otherwise. Such optimal cutoff points depend on the specific structure of the organization as well as on the position that the agent has in the organization. Koh (1992b) shows that optimal acceptance regions are smaller (larger) at lower level of the hierarchy (polyarchy), compared with higher levels. Visser (2000) takes a step further by allowing rational agents to have heterogenous screening accuracies, and finds that the ordering of agents does not affect the performance of the organization.

All models that take human fallibility as a standpoint share the assumption that projects come in streams, the organization screens them sequentially, and eventually undertakes all projects perceived as profitable. However there are situations in which it is more realistic to assume that the task performed by an organization is to select a single alternative from a feasible set of alternatives that are simultaneously available. Think, for instance of the problem faced by a university that has a vacancy to offer. The university will typically post the job offer on the Internet (e.g., on Inomics or the AEA website) and, after screening a number of applicants, it will hire the most promising candidate.

The consideration above motivates the third and fourth chapter of the present thesis, where I take the human fallibility approach as point of departure to analyze the collective decision-making process of an organization whose goal is to select the best alternative from a given set. The analysis is restricted to a hierarchy as it is the most effective organizational structure in performing such task. Each alternative is assumed to be either good or bad, and human fallibility is captured by the fact that each agent observes, for each alternative, a noisy signal in the form of good news or bad news about the alternative's type. With this setup a hierarchy with two levels represent a two-stage selection procedure where the first agent preselects a subsample of alternatives to be passed to the second stage, and the second agent makes the final selection by choosing one of the preselected alternatives. I abstract from both fixed and variable costs (e.g. the cost of hiring screeners and the cost of evaluating alternatives) as it allows me to analyze the organizational performance only in terms of the screening accuracy.

In Chapter 3 I analyze the effects that two different information processing limitations (limited attention and lack of communication) have on the optimal selection strategy by keeping the ordering of agents exogenous. Besides lack of communication and limited attention, I do not impose any other constraints on individual rationality and assume that agents are team members whose goal is to maximize the probability that the organization selects a good candidate. If information processing limitations were absent the decision-making process (besides the imprecision due to human fallibility) would be perfect: the first agent should pass all alternatives to the second stage, and the second agent eventually selects the alternatives that has generated the most favorable pair of signals. I first analyze the collective decision-making process when communication between agents is not possible. Later, I will allow for the possibility of full communication, but I will assume that agents have limited attention: the signals' accuracy decreases in the number of alternatives to be evaluated. Moreover, I allow for agents to be heterogenous in their screening accuracies.

My first finding in chapter 3 is that, when communication is not possible and the more accurate agent is placed at the first decision node, the optimal selection strategy prescribes to pass to the second stage only alternatives that have generated good news. This is the case because, when the more accurate screener acts first, there are no incentives to pass to the second stage alternatives that produced bad signals: independently of the signal realizations in stage 2, they will never produce a likelihood value larger than that of alternatives that produced a good signal in stage 1. When, in contrast, the most accurate screening takes place in the second stage, the optimal selection strategy generally prescribes to pass to the second stage (in addition to all alternatives that have generated good news) also some of the alternatives that have generated bad news. This is the case as passing bad news to the second stage implies a trade-off between two effects. The first is a negative mixing effect: mixing good and bad news in a single pool makes it impossible to distinguish between them in stage 2 and, therefore decreases the probability of selecting the best alternative. The second is a positive sample size effect: selecting more alternatives in stage one increases the chance of observing at least one good news in stage two. My second finding is that, when I allow for the possibility of internal communication but I impose limited attention, it is generally optimal for the first agent not to pass too many alternatives to the next stage. Passing a large sample of alternatives to stage 2 results in overloading the second agent by excessively reducing his screening accuracy. Consequently, the optimal selection rule takes the form of a bound on the total number of alternatives (both good and bad news) to be passed to the next selection round.

In chapter 4 I focus only on lack of communication but I endogenize the ordering of agents. I find that, by letting the more accurate agent screen first, the organization overcomes the imperfections due to lack of communication and is as efficient as a hierarchy in which information flows are perfect. Even though this result suggests that, by an optimal ordering of agents, perfect decision-making can always be achieved, it is important to notice that variable costs of screening have not been taken into account. As it is natural to assume that the cost of processing information increases with the accuracy with which information is processed, it might be the case that having higher screening accuracy at later stages is optimal for the simple reason that at higher stages less information is processed. Consequently, the indirect (e.g. non monetary) costs that might emerge due to a lack of internal communication can always be explained in terms of the trade-off between mixing effect and sample size effect highlighted in chapter 3.

4 Individual decision-making and imperfect information processing

The idea that the traditional economic paradigm of rationality imposes excessively demanding computational abilities to the human brain was first pointed out by Simon (1955). The original critique of Simon has been supported by abudant empirical evidence from the psychological literature that shows how human choice procedures differ from that of *homo economicus.* Tversky and Kahneman (1981,1986) show the existence of framing effects: choices are affected by the way in which logically equivalent alternatives are presented. Tversky and Kahneman (1986) and Tversky (1969,1977) provide evidence suggesting that, when making decisions, humans have the tendency to simplify choice problems rather than analyzing the consequences of choosing each alternative in an exhaustive quantitative manner. Furthermore, experiments by Huber Payne and Puto (1982) and Tversky and Shafir (1992) demonstrate that choices are often based on reasons that are dependent on the specific choice problem. Luce and Raiffa's (1957) dinner is a striking example of this phenomenon. An individual chooses chicken from a menu offering steak and chicken. However the same individual chooses steak when the menu includes steak chicken and frog's legs. The reason explaining this seemingly irrational choice behavior is that the presence of frog's legs in the menu conveys information about the quality of the other two items. Framing effects, tendency to simplify problems and the presence of choice set-dependent reasons, they all lead to choice functions that generally violate the Independence of Irrelevant Alternatives Axiom (IIA) and, therefore, imply a behavior which is not consistent with the rationality paradigm[3]. The tension between the behavior predicted by traditional economics and the behavior observed in experiments, gave rise to several attempts (often very different in their nature) to model bounded rationality[4].

One approach is to impose some constraints on the information processing capacity of the decision-maker. Anderlini (1991), Spear (1989) and Anderlini and Felli (1992) assume that an agent can use only a recursive behavior rule. Rubinstein (1986), Abreu and Rubinstein (1988) and Kalai and Stanford (1988) analyze repeated games under the assumption that more complex strategies are more costly to use. In this setup players are rational but take into account complexity costs when choosing the strategy to implement[5]. Rosenthal (1989), Chen et. al. (1997), and Rubinstein (1998), initiated a study of bounded rationality within the framework of game theory.

Another approach is to assume that the environment in which agents operate is prohibitively complex. For example, Easley and Rustichini (1999, 2005) model the choice behavior of an individual who does not have states of the world as part of his description of the decision problem. Instead the agents know the set of feasible actions, he formulates preferences over these actions, and chooses accordingly to these preferences. The analysis focuses on determining a systematic (expected payoff-maximizing) procedure which describes how the decision-maker modifies his preferences on the basis of experience.

Other authors have undertaken the challenging task of refining the traditional concept of rationality (referred to as substantive rationality) into a rational behavior (procedural rationality) that is compatible with the patterns persistently observed in experimental evidence. One research agenda

[3] To this respect, Sen (1993) argues that "there is no way of determining whether a choice function is consistent or not without referring to something external to choice behavior".

[4] For a survey of this literature see Rubinstein (1988a).

[5] These papers belong to the literature on games and automata: Players are restricted to using finite state automata to implement their strategies. An automaton consists of a set of internal states (one of which is the initial states), a transition function (that specifies how the automata change in response to the opponents' actions) and an output function (which gives an action as a function of the state).

is to model formally choice procedures supported by empirical evidence, and to investigate whether or not such procedures are compatible with the rationality paradigm. In case they are, the task becomes to identify restrictions on the preference space that are compatible with those procedures. For instance, Rubinstein (1988b) formalizes procedural decision-making taking into account the role of similarities in human reasoning which was emphasized by Tversky (1969, 1977). Differently, Gilboa and Schmeidler (1995,1997), model the behavior of a decision-maker who displays case-based reasoning: decisions are based on the consequences derived form past actions taken in similar situations.

The focus of more recent contributions is that of finding plausible ways to rationalize bounded rational choices. Kalai, Rubinstein and Spiegler (2002) argue that choices are not the result of a single rationale, but that of multiple rationales, each one being applied to a subset of choice problems. Their approach allows in principle to rationalize all choices and the focus becomes to count the minimum number of rationales necessary to explain a given choice. Differently, Manzini and Mariotti (2005) define a sequentially rationalizable choice function as a choice which can be retrieved by applying sequentially a fixed set of rationales to remove inferior alternatives. This approach views choice as a sequential elimination heuristics and translates in economic terminology procedures that have already been proposed by psychologists. For instance, it is clear the analogy with Tversky (1972a,1972b) model of "Elimination by Aspects" (EBA): each alternative is represented by a set of aspects, and the choice is made by a sequential elimination of alternatives that do not posses an aspect.

Even though the economic literature has proposed and analyzed several choice procedures, there is, to my knowledge, no work that analyzes in a systematic way the problem of a utility-maximizing decision-maker who has to choose the best out of several feasible procedures. This motivates chapter 5 of this thesis, where I analyze the choice function of a decision-maker who is unable to evaluate his objective function. I take a procedural approach to decision-making and model it as a selection procedure of Choice by Elimination (CBE) (Tversky (1972a,1972b)). This approach assumes that each alternative is viewed as a sequence of aspects that the alternative may or may not posses. In my model each aspect has the interpretation of being a good news about the quality of the corresponding alternative. The decision-maker's inability to evaluate the objective function is captured by the fact that the different aspects (news) of each alternative cannot be aggregated into a single performance measure (i.e., likelihood). With this setup the decision-maker's problem is that of choosing the procedure (e.g., a rule that prescribes, upon observing a set of news – each one corresponding to a single alternative – which alternatives should be eliminated along the way) that maximizes the probability of choosing the best alternative. It is important to stress that the decision-making problem I consider can be interpreted as one with imperfect information processing. With this interpretation, the model can be described as follows. There is a sequence of stages and, in each stage, information is freely revealed to the decision-maker in the form of a set of signals, where each signal imprecisely reveals the type of an alternative. The decision maker's inability to evaluate an objective function is then equivalent to assuming that the decision problem is one with imperfect recall: the set of information is partitioned in different subsets (each one representing the set of signals observed in each stage) that cannot be simultaneously used in order to single out one alternative. I am able to show that the procedure that maximizes the probability of choosing the best alternative is unique and I call it 'Single Worst Elimination' (SWE) as it prescribes to eliminate sequentially only one alternative, and only when such alternative is the singleton bad news. Surprisingly I find that the choice function induced by such procedure is transitive but violates the Weak Axiom of Revealed Preferences.

2. Markets with information overload

1 Introduction

In this chapter I investigate the functioning of a market in which the agent in charge of processing information has limitations in doing so. Information has through the chapter a quantitative interpretation, implying that, a larger amount of information requires more resources in order to be interpreted.

The rise of the Internet had a massive impact on the way economic agents produce, transmit, acquire and process information. Nowadays an employer can post a job opening with fixed terms on a specific web page and potential employees can simply react with a "click" by sending an electronic file that includes, for instance, a CV, or a letter of recommendation. To make things more concrete, think of the role that the web pages of the AEA or Inomics play in the academic labor market. Similarly think of how the possibility of sending a paper in electronic format has effected the market of scientific publications. At least two characteristics distinguish theses types of markets from more traditional ones. The first one is the potentially unlimited number of applicants (i.e., all potential candidates that have access to the Internet). The second is the relatively low cost at which information can be transmitted (i.e., the cost of sending a file via e-mail). For markets in which there is such an abundance of information it is fairly realistic to think that the scarce resource has become the human attention needed to process and understand such information. For example, the cost of sending an electronic manuscript is much lower than the opportunity cost of the time it takes a referee to read and understand the paper. These considerations motivate the present chapter where I investigate a monopsonistic labor market in which the monopsonist has limited attention to process the whole information he receives.

I develop a model in which heterogenous senders (job applicants) compete in order to be selected by one receiver (employer). Productivity is private information to the senders, and the receiver processes imprecise informative signals about workers' productivity (applications) in order to screen among applicants. I abstract from analyzing wage mechanisms because they have already been studied in screening models without information processing limitations[1], and because this is not essential for my analysis. The informational asymmetry is captured by the fact each sender has private

[1] See for instance the recent work by Janssen (2002) and Delfgaauw and Dur (2005).

information about his own productivity but is not fully informed about how his application will be perceived by the receiver, whereas the receiver, by processing an application, can only observe a noisy signal about the unknown productivity of the corresponding applicant. Limited attention is modeled by assuming that the information processing technology is imperfect: the accuracy of each signal in predicting the unknown productivity decreases with the total number of applications processed.

My first finding is that, when cost of sending application is low relatively to the existing technology level, some information is disregerded by the receiver in equilibrium. Moreover, the lower the sending cost, the larger is the amount of information disregarded. The receiver's ability to process an optimal amount of information may suggest the idea that receiving information that exceeds the processing capacity does not necessarily mean receiving too much information. However information overload is present in my settings as I am also able to show that the larger is the amount of information disregarded in equilibrium, the lower is the average quality of the overall information received. This result is confirms the analytical results of Van Zandt (2004).

Van Zandt (2004) addresses information overload issues for a network of targeted communication: several firms (senders) send advertisement messages to an arbitrary subset of consumers (receivers). Limited attention is modelled by assuming that receivers can process only a fixed amount of messages: if they receive more messages than they can process, they randomly select a number of them without knowing their content. The main result of Van Zandt (2004) is that information overload is a consequence of the strategic interaction among senders. Since the attention of receivers is an unpriced resource, each sender's message crowds out messages of other senders and an increase in the amount of information results in a decrease in its average value.

Besides the analogy between my result and that of Van Zandt, the mechanism shaping it are very different. In Van Zandt (2004) senders are competing only for the attention of the receiver, and the sending cost is a rationing device of receivers' attention: a higher cost induces senders to screen more effectively the receivers to target. Differently, the senders of my model compete in two dimensions. First, there is a competition for attention because a necessary condition for a sender to be offered the vacancy is that his application is taken into account in the first place. Second, all senders whose applications have been processed compete in terms of their perceived productivity because the vacancy is eventually offered to the sender whose application has been perceived by the receiver as being the best one. Consequently, the cost of sending application works in my model as a screening device which induces only the more productive senders to apply. Further, contrary to Van Zandt (2004), where disregarding information is a behavior exogenously imposed by assuming a hard bound on the number of messages that can be processed, in my setup, the receiver can choose explicitly now much information to use. This allows me to model more accurately the receiver's decision of how much attention to invest in processing information. The optimal amount of information to process is determined by comparing the marginal benefit of a larger sample size (which is the chance of observing a better application) to its opportunity cost (which is the marginal decrease in the capability of ranking applicants according to their true abilities). The nature of such optimal behavior is also supported by findings in the experimental marketing literature. Jacoby et al. (1974) find that the accuracy which consumers make a right purchase is a bell-shaped function of the amount of information at disposal: too little as well as too much information results in a poor decision, implying that, in principle, there exists an optimal intermediate level of information to process.

Finally, I share with Van Zandt (2004) the idea that signals are unstructured: (a) the receiver cannot distinguish between signals when choosing which ones to process, and (b) each signal is either processed in its entirety or not at all. This assumption precludes the possibility of sequential screening based on partial processing of signals. For instance, if CV's were meant as unstructured signals, a first screening of applicants could be based only on the specific field in which each candidate works, a second screening could be based on the number of publications, a third on the knowledge of foreign languages, and so on. Such screening methods are important mechanisms to deal with information overload, however certain situations imply necessarily a simultaneous screening of information. A typical example is the parallel infromation-processing that takes place in organizations, where agents have a limited time to screen a predetermined batch of information.

The second issue I address is the effect of market size on information overload. Surprisingly, I find that, if the size of the market (that is, the number of senders) increases, the amount of information neglected in equilibrium stays constant and the average quality of information increases. This occurs because of the self-selection mechanism that typically operates in screening markets: the increase in competition discourages low-productive senders from applying and, therefore, a larger market translates only in higher average productivity of applicants.

Finally, I note that the assumptions that information is a freely disposable good might be too restrictive. First, Jacoby et al. (1974), Malhotra (1982), and Keller and Staelin (1987), provides empirical evidence that consumers are not able to shield themselves from being overloaded when too much information is made readily available[2]. Second, there are situations in which legal constraints or specific policy orientations preclude the possibility of neglecting information. For instance, it is common for referred journal

[2]For a methodological critique to this findings see, in contrast, Meyer and Johnson (1989).

to process all the manuscripts received. Therefore, in the last section of the chapter I assume that the receiver must process the whole information received. With this setup, information overload refers to the difference between the total number of applications sent and the number of applications that would be optimal to process. I find that the inefficiencies of information overload are now even more severe. There are two effects responsible for this result. The first is trivial and is due to the fact that excessive information directly harms the receiver who, contrary to the case in which information is freely disposable, cannot shield himself form being overloaded. The other effect originate from the senders' side. I am able to show that, for any given technology level, market size, and (sufficiently low) sending cost, inability of disregarding information results in a larger amount information being sent as well as lower quality of such information. This is so because the competition between senders is weakened. First, senders do not compete anymore for attention as they know that all applications sent will be processed. Second, the screening technology of the receiver is used beyond its optimal processing capacity and, thus, sorting out relatively bad candidates becomes more difficult.

The chapter is organized as follows. Section 2 describes the model. Sections 3 and 4 analyze the problems of the Receiver and that of the senders. Section 5 provides the analysis. Section 6 investigates the relation between market congestion and inefficiency. Section 7 concludes the chapter and section 8 includes the proofs.

2 The model

In this section I first describe the model and discuss the related economic implications. I then provide and justify a more specific assumption on the information processing technology.

There is a mass N of heterogenous potential senders, where N is a real number $1 \leq N < \infty$ and a mass, normalized to one, of homogenous receivers. Each receiver has a vacancy to offer to a sender. The mass of senders is uniformly distributed on the $[0,1]$ segment. Each point on the segment corresponds to the ability of a sender and is denoted by $\theta \in [0,1]$. The decision variable for a sender is whether to send a costly application or not. The decision variables for a receiver are: how many applications to process, and which application to select from the pool of those processed. I denote by $n \geq 0$ the mass of senders who actually apply, and by m the mass of senders actually processed by the receivers. Clearly $0 \leq m \leq n \leq N$. The timing is as follows.

0. Each sender knows where he is located (he knows his own ability), and decides whether or not to send an application. I denote the sender's choice variable by s, where s can take only two values: 0 (do not send) and

1 (send). The cost of sending one application is the same for all senders and is equal to C. I will call *actual* senders those senders who have applied.

1. A random matching function allocates the mass of actual senders n, evenly to the unitary mass of receivers. Thus, each receiver receives a fraction n, of applications.

2. Each receiver observes the fraction of applications at his disposal, n, and decides how many of them to process, m. Therefore the strategy for a receiver is denoted by $m(n)$.

3. Let $\Theta = \{\theta : s(\theta) = 1\}$ be the set of actual senders, let $F(\theta \mid \theta \in \Theta)$ be the prior distribution of actual senders, and let X be a subset of $\mathbb{R}$ representing the set of possible signals about θ. The receiver simultaneously processes m applications. When an application is processed it produces a signal which is drawn from the distribution $F(\cdot \mid \theta, m)$[3]. It is assumed that signals draws are i.i.d., and the family of distributions $\{F(\cdot \mid \theta, m)\}$ is such that, for any $x_2 > x_1$ the posterior distribution $G(\cdot \mid x_2, m)$ dominates the posterior distribution $G(\cdot \mid x_1, m)$ in the sense of strict first-order stochastic dominance. The crucial assumption of the model is the idea that, when infinitely many signals are processed, then each signal becomes completely uninformative. This is captured by the following assumption: for any $\theta \in \Theta$ and $x \in X$

$$\lim_{m \to \infty} F(x \mid \theta, m) = F(\theta \mid \theta \in \Theta) \qquad (A.1)$$

4. The set of signal outcomes, denoted by $\hat{X}_m \equiv \{\hat{x}_1, .., \hat{x}_m\}$, becomes private information to each receiver who compares signal outcomes and selects the one that maximizes his expected utility (that is, the highest signal observed).

The payoffs are as follows. If sender θ does not apply he gets utility zero. If sender θ applies then the utility he gets equals the value of the vacancy (constant for all senders and denoted by V, where $V > C$) weighted by the probability of being selected minus the cost of signalling. The probability that sender θ assigns to the event of being selected will generally depend on his own ability, the total number of applications sent, and the number of applications actually processed and, hereafter, will be denoted by $\Phi(\theta \mid n, m)$.Therefore, if n senders apply, and m are processed, the payoff of sender θ is

$$u^s(\theta, n, m) = \begin{cases} 0 & \text{if } \; s(\theta) = 0 \\ \Phi(\theta \mid n, m) V - C & \text{if } \; s(\theta) = 1 \end{cases} \qquad (1)$$

The payoff that a receiver gets from selecting the sender whose signal outcome is $\hat{x}$ equals the expected ability of that sender. The expected ability of sender θ is conditional on the value of the corresponding signal outcome, $\hat{x}$, and the total number of applications processed m. Thus, the ability of a

[3] In the analysis section, however, I will focus on a more specific distribution.

sender θ, as it is perceived by the receiver, is

$$u^r(\hat{x}, m) = E_m[\theta \mid \hat{x}] \tag{2}$$

The assumption of stage 1 is justified by the purpose of this chapter which is not that of explaining the mechanism by which senders are assigned to receivers. Since receivers are homogenous, a sender does not perceive a receiver per-se more valuable than another. Therefore I intentionally avoid to model the possibility for a sender to target a particular receiver. The only crucial choice of a sender is that of sending an application or not, where sending an application is indeed a choice of whether to enter the market. When applications are sent, the size of the actual market, (the mass of actual senders n), is endogenously determined and stage 1 provides a stylized picture of the way in which the two sides of the market come to meet each other. The unitary mass of receivers corresponds to a mass of identical information processors, thus each receiver can be thought as the average representative agent of the receivers' population. From now on I will use the term Receiver (with capital R) to mean the representative receiver. Stage 1 also implies that n is not an integer, thus, it does not represent the number of applications received but is a measure of the information produced by the market. This is convenient for the analysis and, with abuse of terminology I will call n simply the "number" of applications received.

An important feature of the model I propose is that information is revealed to the Receiver by observing signals that are drawn simultaneously. This assumption describes economic situations in which a decision maker decides ex-ante how much to invest in the information processing phase and, only after such decision is made, the informative outcome is revealed[4].

The assumptions of stage 3 imply that signals are imperfectly informative: higher signal outcomes are *more favorable than*[5] lower signal outcomes. Moreover, assumption $(A.1)$ implies that when infinitely many applications are processed the infromativeness of each signal vanishes as each signal outcome becomes *equivalent* and *neutral*[6] to the decision maker. This assumption is economically justified as one can easily think of situations in which the agent in charge of processing information has limited resources in order to accomplish this task. A direct consequence of such constraint is that, if the Receiver processes infinitely many applications, he actually allocates

[4] To this respect, Moscarini and Smith (2001) show that, in a dynamic continuous time world, one shot non sequential sampling is still optimal given discounting and a constant marginal cost of information.

[5] The concept of "*favorableness*" was first introduced by Milgrom (1981). If Θ is the set representing the possible values of the random parameter θ, and X is the set of possible signals about θ, then, a signal x_2 is *more favorable than* a signal x_1 iff the posterior distribution $G(\theta \mid x_2)$ dominates in the first-order stochastic sense the distribution $G(\theta \mid x_1)$.

[6] Two signals x_1 and x_2 are *equivalent* if $E(\theta \mid x_1) = E(\theta \mid x_2)$. A signal x is *neutral* if, for any prior distribution $G(\cdot)$, $G(\theta) = G(\theta \mid x)$.

zero resources to interpret each one of them and signals turn out to be completely uninformative.

The cost of applying is assumed not to be related to the senders' ability since its interpretation is not that of an opportunity cost, but simply the mere cost of transmitting information (i.e., filling in and sending one application). On the other hand, in this model, the opportunity cost is captured by the fact that more able senders have higher probability of being selected.

It is useful to clarify that both asymmetric information and imperfect information processing technology are crucial to the model. On one hand, when dealing with the problem of selecting the "best" sender, the Receiver can only make an inference on the senders' ability by using a noisy signal that imprecisely represents the true ability. On the other hand, when dealing with the problem of inferring his chance of being selected, a sender can only use the knowledge about his ability to predict what signal outcome his application will produce. In the model there is a discrepancy between the value of a signal outcome and the true ability value that the signal aims to represent and, moreover, such discrepancy increases with the total number of applications processed. Therefore the larger is the amount of information processed by the market, the less is the efficiency with which agents can use their private information in order to make a decision.

Finally, the model can also represent a situation in which neglecting applications is not an option. This can be done by restricting the strategy space of the Receiver to be simply that of selecting one applications form those received. These two scenarios, one in which applications can be neglected, and one in which this is not possible, are compared in section 6.

2.1 The information processing technology

In this subsection I propose a more specific information processing technology which has two desirable properties. Namely I will impose that the two following requirements are met:

(i) The information processing technology implies a *trade-off between the quality and the quantity of information.*

(ii) The information processing is *not sensitive to the ability of actual senders.*

Referring to property (i), by quality of information I mean the accuracy of each signal in predicting the unknown parameter. By quantity of information I mean the total number of signal outcomes drawn simultaneously[7]. Then the first property requires that the signals' accuracy is strictly decreasing in the sample size; the economic interpretation is that, the less

[7] The same definition of quantity of information is provided in Moscarini and Smith (2002).

resources are allocated to interpret each signal, the smaller is the accuracy of each signal in predicting the unknown parameter.

Property (ii) ensures that the signals' accuracy depends only on the total number of signals drawn and not on the specific set of parameter that signals aim to represent. This property plays an important role as the set of possible ability values of actual senders may be different for different equilibria. Indeed, notice that if a sender with ability, say, $\bar{\theta}$ applies also senders with ability above $\bar{\theta}$ apply. This allows for the existence of a partially separating equilibrium in which the segment of actual senders is $[\alpha, 1]$, where α denotes the ability of the sender who is indifferent between sending or not sending an application. Therefore the senders' strategic behavior affects the set of possible ability values, which, in case of a partially separating equilibrium, is $\Theta_\alpha = [\alpha, 1]$. Consequently, the support of family of distributions from which signal outcomes are drawn also depends on α, and I will henceforth denote it by $\{F_\alpha(\cdot \mid \theta, m)\}$. Property (ii) ensures that the only effect that α has on the distribution $F_\alpha(\cdot \mid \cdot)$ is that of rescaling it according to the ability support of actual senders, $[\alpha, 1]$.

In the following a more specific information processing technology, which has both properties (i) and (ii), is introduced.

$(A.2)$ Let $\Theta_\alpha = [\alpha, 1]$ be the set of possible ability values, where $\alpha \in [0, 1)$. Then, in stage 3 signal outcomes are drawn from the family of distributions $\{F_\alpha(\cdot \mid \theta, m)\}$, where

$$F_\alpha(x \mid \theta, m) = I_{[\alpha,1]}(x)(1 - \pi(m))\left(\frac{x-\alpha}{1-\alpha}\right) + I_{[\theta,1]}(x)\,\pi(m) + I_{(1,\infty)}(x) \quad (3)$$

and where $\pi(\cdot)$ is a twice differentiable function such that $\pi(1) \equiv \pi_1 \leq 1$, $\pi'(\cdot) < 0$, $\pi''(\cdot) > 0$ and $\lim_{m\to\infty} \pi(m) = 0$.

In the following pictures the distribution $F_\alpha(\cdot \mid \theta, m)$ and the corresponding density function $f_\alpha(\cdot \mid \theta, m)$ are depicted for the case in which $\alpha = 0$.

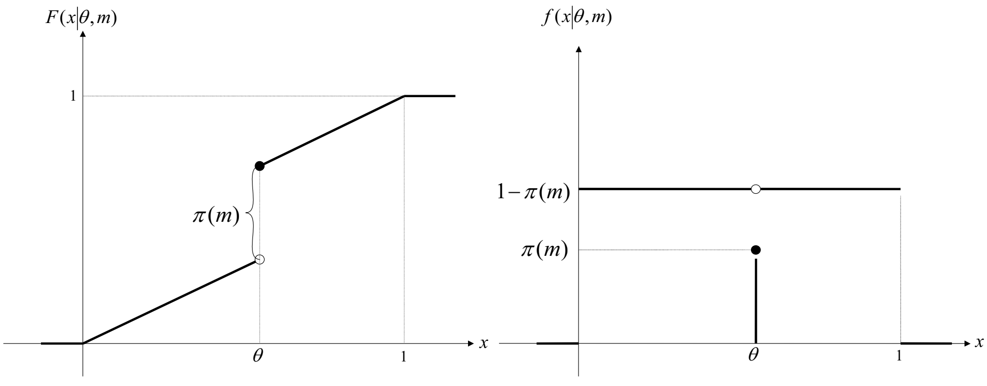


Figure 1(a) Figure 1(b)

$F_{\alpha}(\cdot \mid \theta, m)$ is a mixture distribution: the distribution of the signal conditional on ability θ is, with probability $\pi(m)$, a degenerate distribution with all probability mass at $x = \theta$, and, with probability $1 - \pi(m)$, a uniform distribution on $[\alpha, 1]$. Thus, $\pi(\cdot) \in (0, \pi_1]$ is the revealing probability and denotes the accuracy of each signal in predicting the unknown ability: the higher the value of $\pi(\cdot)$ the higher the accuracy, and $\pi(\cdot) \to 0$ captures the situation in which signals become completely uninformative. Since $\pi(\cdot)$ is decreasing in m, eventually approaching zero, the signals' accuracy, vanishes with the number of signals that the Receiver processes. To see it, notice that, if $m \to \infty$, then $\pi(m) \to 0$ and $f_{\alpha}(x \mid \theta, m) \sim I_{[\alpha,1]}(x) \frac{1}{1-\alpha}$ for any θ. In other words, as $m \to \infty$, all signals become *equivalent* and *neutral* to the decision maker. It is the monotonic behavior of $\pi(\cdot)$ which captures the *trade-off between the quality and the quantity of information.* The fact that the information processing is *not sensitive to the ability of actual senders* is captured by the fact that the only effect of α on the distribution $F_{\alpha}(\cdot \mid \cdot)$ is that of rescaling it according to the ability support of actual senders, $[\alpha, 1]$. The implications of such property will be discussed at the end of section 3 and section 4. In the remaining of the chapter it is assumed that the information processing technology is the one defined by $(A.2)$.

3 The problem of the Receiver

Since the payoff of the Receiver depends on the support of actual senders, $\Theta_{\alpha} = [\alpha, 1]$, henceforth, it will be denoted by $u_{\alpha}^{r}(\hat{x}, m) \equiv E[\theta \mid \hat{x}, m, \theta \geq \alpha]$. The following lemma characterizes the functional form of the terminal payoff of the Receiver.

Lemma 2.1. *The payoff of the Receiver is*

$$u_{\alpha}^{r}(\hat{x}, m) = \pi(m)\hat{x} + (1 - \pi(m))\left(\frac{\alpha + 1}{2}\right) \quad (4)$$

The interpretation of (4) is straightforward. The information perspective of the Receiver is a mirror image of that of the sender, as he observes x but not θ, moreover, $\pi(\cdot)$ and $1 - \pi(\cdot)$ represent respectively the probability that a signal is fully informative and the probability that a signal is completely uninformative. Therefore, the Receiver perceives the ability of a sender as being equal to the corresponding signal outcome, with probability $\pi(\cdot)$, and as being equal to the prior average ability of actual senders with probability $1 - \pi(\cdot)$. Notice also that (4), considered as a function of signal x, is a straight line with slope $\pi(m)$. Thus, the more signals are processed, the flatter is $u^{r}(\hat{x}, m)$; this behavior captures the fact that the informativeness of each signal decreases as more signals are simultaneously processed. Finally, the fact that $\lim_{m\to\infty} u^{r}(\hat{x}, m) = (\alpha + 1)/2$ is the feedback on the

Receiver's side of the gradual degeneration of the signals' informativeness. In the following the optimal strategy of the Receiver is determined.

Proposition 2.1. *The dominant strategy for the Receiver is* $\{m^*(n), \hat{x}^*\}$, *where* $\hat{x}^* = \max\{\hat{x}_1, .., \hat{x}_m\}$ *and* $m^*(n) = \min\{n, \bar{m}\}$, *with* $\bar{m} \in (1, \infty)$. *Moreover,* $\bar{m}$ *is unique and is given by the following condition*

$$\frac{|\pi'(\bar{m})|}{\pi(\bar{m})} = \frac{2}{(\bar{m}^2 - 1)} \tag{5}$$

I will provide a sketch of the proof of proposition 2.1 followed by some general remarks. The Receiver solves the following maximization problem

$$\max_{m,\hat{x}} u_\alpha^r(\hat{x}, m)$$

$$\text{subject to} \quad : \quad \begin{cases} m \leq n \\ \hat{x} \in \hat{X}_m \end{cases}$$

He chooses the sample size m in stage 2 and selects one out of m signal outcomes in stage 4. Consider first stage 4. Since $u_\alpha^r(\hat{x}, m)$ is strictly increasing in $\hat{x}$, no matters how many (finite) signals the Receiver processes, he always selects the sender whose application produced the highest signal outcome. This proves $\hat{x}^* = \max\{\hat{x}_1, .., \hat{x}_m\}$. Denote by $\bar{x}_{(m)} \equiv E_x[\max\{x_1, ..x_m\}]$ the expected value of the maximum of m signals, and consider now stage 2, when the receiver chooses m in order to maximize the ex-ante expected utility

$$\begin{aligned} E_x[u_\alpha^r(x^*, m)] &= u_\alpha^r\left(\bar{x}_{(m)}, m\right) \\ &= \pi(m)\,\bar{x}_{(m)} + (1 - \pi(m))\left(\frac{\alpha + 1}{2}\right) \end{aligned} \tag{6}$$

Notice first that, $u_\alpha^r\left(\bar{x}_{(1)}, 1\right) = u_\alpha^r((\alpha + 1)/2, 1) = (\alpha + 1)/2$, that is, processing only one signal implies selecting one sender randomly and, therefore, getting ex-ante payoff equal to the prior expected ability of actual senders. Moreover, $u_\alpha^r\left(\bar{x}_{(m)}, m\right)$ approaches $(\alpha + 1)/2$ as $m \to \infty$; this is the case because, when infinitely many signals are processed, each signal becomes completely uninformative. This implies that the ideal number of signals to process, namely $\bar{m}$, is always finite. Moreover, the well-behavior of function $\pi(\cdot)$ ensures that such number is unique and I will now provide the intuition behind its characterization. The function $u_\alpha^r\left(\bar{x}_{(m)}, m\right)$ captures the trade-off between quality and quantity of information that the Receiver faces when he has to choose m. The argument $\bar{x}_{(m)}$ increases in m, that is, the more signals are processed the higher is the chance of observing a very high signal. This is the *sample size* effect which has, ceteris paribus, a positive impact on the terminal payoff. However, the Receiver can increase the chance of observing a high signal only at the cost of reducing its quality.

This is the detrimental effect of *decreasing accuracy* which is captured by the fact that $u^r_\alpha(\cdot, m)$ becomes flatter the larger is m (see fig. 2(a)). Therefore the ideal number of signals to process, $\bar{m}$, is determined by the balancement of the two effects. Condition (5), which characterizes $\bar{m}$, has a clear cost benefit interpretation since it is obtained by equating the marginal benefit of processing one additional application to the marginal cost of doing so. Notice from (6) that

$$\begin{aligned} MB(m) &= \pi(m) \frac{d\left(\bar{x}_{(m)}\right)}{dm} \\ MC(m) &= \left|\pi'(m)\right| \left(\bar{x}_{(m)} - \frac{\alpha+1}{2}\right) \end{aligned}$$

that is: the marginal benefit of processing one additional application is the marginal increase of the expected value of the maximum signal, weighted by the accuracy of the signal; the marginal cost of processing one additional application is the (positive) difference between the expected maximum signal and the prior average ability, weighted by the marginal decrease of the signal's accuracy. From the discussion above it follows that $u^r_\alpha\left(\bar{x}_{(m)}, m\right)$, considered as a function of m, is bell-shaped and reaches its unique maximum at $\bar{m} \in (0, \infty)$ (see figure 2(b)).

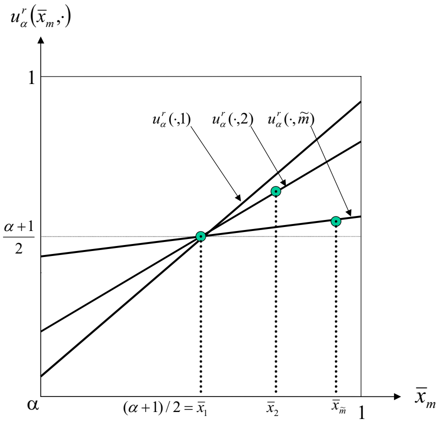


Figure 2(a)

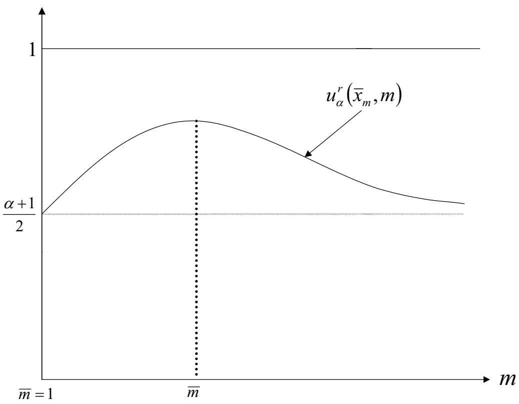


Figure 2(b)

Notice that $\bar{m}$ denotes the number of applications that the Receiver is, at most, willing to process. This implies that, if the number of applications received, n, is larger than $\bar{m}$ the Receiver is overloaded, as $n - \bar{m}$ applications are simply neglected as a result of the maximizing behavior.

Two aspects of proposition 2.1 are worth noticing. First, the optimal behavior of the Receiver does not depend on α: this is a direct consequence of the fact that the *information processing technology is not sensitive to the ability of actual senders*. Second, $\bar{m}$ is fully determined by condition (5), which follows directly from assumption $(A.2)$. However, it is important to stress that, in order to allow for the possibility of information overload, assumption $(A.1)$ is sufficient. To see it notice that as m becomes large the Receiver views all signal as *equivalent* and *neutral*, and therefore each signal yields expected utility equal to the prior average ability. Consequently also under $(A.1)$ the optimal number of signals to process is finite. Therefore the only convenience of assumption $(A.2)$ is that it allows to have a closed-form solution to the maximization problem of the Receiver which does not depend on α and which has a clear cost-benefit interpretation.

4 The problem of the sender

If $\Theta_\alpha = [\alpha, 1]$, when n senders apply and m applications are processed, the expected utility of an arbitrary sender $\theta \in [\alpha, 1]$ is

$$u^s_\alpha(\theta, n, m) = \Phi_\alpha(\theta \mid n, m) V - C$$

where $\Phi_\alpha(\theta \mid n, m) \equiv \Phi(\theta \mid n, m, \theta \geq \alpha)$. Denote by $c = C/V$ the cost of applying relative to the value of the vacancy, then sender θ finds it optimal to send if and only if $\Phi_\alpha(\theta \mid n, m) \geq c$. In order to obtain the vacancy, sender θ must go through a two-steps procedure. First, the application of θ must be taken into account by the Receiver, second, provided that this is the case, the signal outcome of θ must be larger than the outcomes of all other applications processed. It is therefore useful to express $\Phi_\alpha(\theta \mid n, m)$ in the following way

$$\Phi_\alpha(\theta \mid n, m) = \gamma(n, m)\, \phi_\alpha(\theta, m) \tag{7}$$

where $\gamma(n, m)$ denotes the probability that one's application is taken into account in the first place when n senders apply and m applications are actually processed, and $\phi_\alpha(\theta, m)$ is the probability that the signal outcome of sender θ is larger than the signal outcomes of the other $m-1$ senders processed. Trivially,

$$\gamma(n, m) = \begin{cases} 1 & \text{if } n \leq m \\ m/n & \text{if } n > m \end{cases} \tag{8}$$

while the nature of $\phi_\alpha(\theta, m)$, is made clear by the following lemma

Lemma 2.2. *If $m \geq 1$ senders are processed by the Receiver, the probability that the signal outcome of sender $\theta \in [\alpha, 1]$ is the largest one is*

$$\phi_\alpha(\theta, m) = \pi(m) \left(\frac{\theta - \alpha}{1 - \alpha}\right)^{m-1} + (1 - \pi(m)) \frac{1}{m} \tag{9}$$

Expression (9) has a clear analogy to expression (4). The first term of (9) says that, with probability $\pi(\cdot)$, the signal is a true representation of ability and, therefore the probability that sender θ assigns to the event of his signal being larger than that of the other $m-1$ senders is simply the $\Pr(\theta \geq \theta_1; ..; \theta \geq \theta_{m-1})$. The second term of (9) states that, with probability $1-\pi(\cdot)$, each signal, independently of the underlying ability, is uniformly distributed on $[\alpha, 1]$; thus, each signal has equal chance of being larger than the other. More importantly, from the expression of $\phi_\alpha(\theta, m)$ it is also clear what is the effect of the decreasing signals' accuracy on the senders' side: as $m \to \infty$, $\phi_\alpha(\theta, m)$ converges in probability to $1/m$, thus, for large samples, the probability of having the largest signal outcome is purely determined by chance. It is important to stress that this phenomenon also holds under the weaker assumption $(A.1)$. If the Receiver allocates his finite resources on infinitely many signals, all signals become *equivalent*. Consequently, the Receiver is not able to discriminate between different signal outcomes and, thus, he randomly selects one of them.

One last property of $\phi_\alpha(\theta, m)$ is worth noticing. Consider the transformation $g(\theta, \alpha) = \alpha + (1-\alpha)\theta$ whose role is that of mapping any $\theta \in [0, 1]$ to a new support, $[\alpha, 1]$, by keeping the relative location of θ fixed. For example, consider the location of the sender with the lowest ability on $[0, 1]$, that is $\theta = 0$; then plugging $\theta = 0$ into $g(\cdot, \alpha)$ yields the location of the sender with the lowest ability on the new support $[\alpha, 1]$, that is, $g(0, \alpha) = \alpha$. Similarly, consider the location of the median sender when the actual ability support is $[0, 1]$, that is $\theta = 1/2$; then plugging $\theta = 1/2$ into $g(\cdot, \alpha)$ yields $g\left(\frac{1}{2}, \alpha\right) = (\alpha + 1)/2$, which is the location of the median sender when the actual ability support is $[\alpha, 1]$. Notice form (9) that it is always the case that $\phi_{\alpha=0}(\theta \mid m) = \phi_\alpha(g(\theta, \alpha) \mid m)$. Such property is a direct implication of the fact that the *information processing technology is not sensitive to the ability of applying senders* and, basically, states that, for a given m, the probability that a sender assigns to the event of his signal outcome being larger than that of an arbitrary opponent, depends only on his relative location on the ability support and not on the particular support chosen.

5 Equilibrium analysis

In the first part of this section I will characterize the equilibria that emerge for specific values of the primitives of the model. In the second part I will provide a comparative static analysis. In order to be able to compare equilibria for different levels of the information processing technology I assume a specific functional form of $\pi(\cdot)$ which depends also on a technology parameter k.

Assumption $(A.3)$

$$\pi(m,k)=\frac{k}{k+m},\text{ where } k\in(0,\infty) \tag{10}$$

The parameter k denotes how good is the information processing technology, and higher values of k imply a better technology. The main convenience of (10) is that k can be interpreted as the amount of resources available to process information, and different values of k allow to capture the entire range of possible information technology levels. For any m, $\lim_{k\to 0}\pi(m,k)=0$, thus, if no resources are available, the information processing technology is completely useless. On the contrary, $\lim_{k\to\infty}\pi(m,k)\to 1$ for any m, thus, if infinitely many resources are at disposal, the technology is perfect. Moreover, the function (10) is increasing in k, meaning that, the more resources available, the more is the information that can be extracted from a fixed number of signals drawn. Now $\pi(\cdot)$ depends also on k, thus, the maximum number of applications that the Receiver is willing to process, $\bar{m}$, will also depend on k and, henceforth, will be denoted by $\bar{m}(k)$. Since a better technology allows to extract more information from the same number of signals it trivially follows that, $\bar{m}'(k)>0$.

As already noticed, more able senders have a higher probability of being selected. Then, since the cost of sending (normalized to the value of the vacancy) c is constant there are two types of equilibria that can emerge: a pooling equilibrium (PE) and a partially separating equilibrium (PSE). The following proposition characterizes the equilibria of the model.

Proposition 2.2. *Let $\pi(\cdot)$ be given by (10). Let be $N>\bar{m}(k)$ and define $\mu(k)\equiv 1/(k+\bar{m}(k))$ and $\rho(k,N)\equiv(\bar{m}(k)/N)\,\mu(k)$. Then:*

$(P.2.2.a)$ *If $c\in[0,\rho(k,N))$, a PE with neglected applications arises.*

$(P.2.2.b)$ *If $c\in[\rho(k,N),\mu(k))$, a unique PSE with neglected applications arises.*

$(P.2.2.c)$ *If $c\in[\mu(k),1/(1+k))$, a unique PSE without neglected applications arises.*

$(P.2.2.d)$ *If $c\in[1/(1+k),1]$ only one sender with expected ability $1-2/N$ applies.*

First, the inequality $N>\bar{m}(k)$ ensures that the market size is large enough for information overload to be a potential problem. If, on the contrary, it was $N\leq\bar{m}(k)$ then, even if the entire mass of senders, N, applies, the information processing technology would never be at full capacity and an equilibrium with information overload would never emerge.

The proposition has a clear graphical interpretation. Let α be the ability threshold level of a PSE. For any PSE, that is, for any threshold level α, it is possible to calculate (see the proof of proposition 2.2) the probability that a sender located at α assigns to the event of being selected in equilibrium. Such probability depends on the primitive of the model N and k and is denoted by $\Psi(\alpha, N, k)$. The qualitative behavior of $\Psi(\cdot, N, k)$, where $N > \bar{m}(k)$, is shown in figure 3

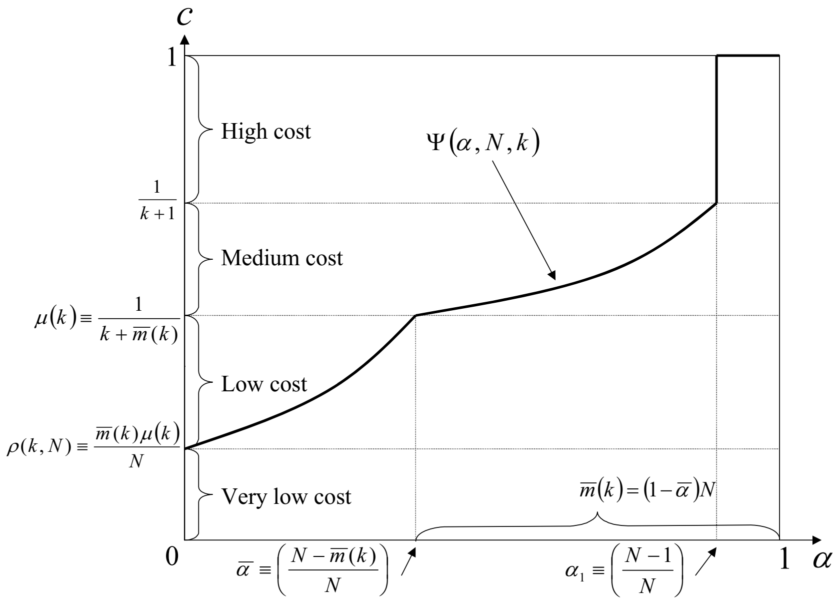


Figure 3

The strict monotonicity of $\Psi(\cdot, N, k)$ is crucial as it implies that a PSE for which the threshold level is α is determined by the condition $\Psi(\alpha, N, k) = c$, while the PE is determined by the condition $\Psi(0, N, k) > c$. Thus, the function $\Psi(\cdot, N, k)$ defines the cost regions for which qualitatively different equilibria emerge and, for any threshold level α, the number of actual senders endogenously determined in equilibrium is given by $n = (1 - \alpha) N$. Then, notice that $\bar{\alpha}$ is the threshold level for which the number of senders applying in equilibrium equals the maximum number of applications that the Receiver is willing to process; indeed, $(1 - \bar{\alpha}) N = \bar{m}(k)$. Therefore, $\mu(k) \equiv 1/(k + \bar{m}(k))$ is the probability that the sender located at $\bar{\alpha}$ assigns to the event of being selected in a PSE for which the threshold level is actually $\bar{\alpha}$. Indeed,

$$\begin{aligned}\Psi(\bar{\alpha}, N, k) &= \phi_{\bar{\alpha}}(\bar{\alpha}, \bar{m}(k)) \\ &= \frac{1 - \pi(\bar{m}(k), k)}{\bar{m}(k)} = \frac{1}{k + \bar{m}(k)}\end{aligned}$$

Similarly, $\rho(k,N) \equiv (\bar{m}(k)/N)\,\mu(k)$ is the probability that the sender located at $\alpha = 0$ is selected in a pooling equilibrium as

$$\begin{aligned}\Psi(0,N,k) &= \frac{\bar{m}(k)}{N}\phi_0(0,\bar{m}(k)) \\ &= \frac{\bar{m}(k)}{N}\frac{1-\pi(\bar{m}(k),k)}{\bar{m}(k)} = (\bar{m}(k)/N)\,\mu(k)\end{aligned}$$

Finally, since $(1-\alpha_1)N = 1$, α_1 denotes the threshold level for which the mass of actual senders in equilibrium is unitary. When this is the case the random matching function ensures that each sender is assigned to one receiver and, therefore, each sender knows that will be selected for sure. This is the reason why, on the one hand $\Psi(\alpha,N,k) = 1$ for any $\alpha \geq \alpha_1$ and, on the other, the Receiver, by selecting the only applicant at his disposal, gets expected utility $(\alpha_1+1)/2 = 1 - 2/N$.

The following corollary follows directly form proposition 2.2.

Corollary 2.1. *For any $k \in (0,\infty)$ and $N > \bar{m}(k)$, a market is overloaded iff $c < \mu(k)$. Moreover $\mu'(k) < 0$ and $\lim_{k\to\infty}\mu(k) = 0$.*

This corollary supports the intuition that market congestion emerges whenever the cost of transmitting information is low, relatively to available resources needed to interpret information. The better the technology, the lower the cost must be, in order for information overload to emerge.

I now study how a marginal change in the market size, N, and in the technology level, k, affects the threshold level α and the number of senders applying in equilibrium, n. To avoid trivialities the attention is restricted to the PSE only. In a PE the sender located at zero is strictly better-off by sending, thus a marginal change in the exogenous parameters will not affect his behavior.

Proposition 2.3. *Let α and n be the equilibrium threshold level and the number of senders applying in a PSE. Then,*

($P.2.3.a$) $\partial\alpha/\partial k > 0$ and $\partial n/\partial k < 0$.

($P.2.3.b$) $\partial\alpha/\partial N > 0$ and $\partial n/\partial N = 0$.

The intuition behind ($P.2.3.a$) is not surprising: a better technology allows the Receiver to rank applicants' quality more precisely, which discourages low-quality employees from applying. On the contrary, the result of ($P.2.3.b$) is very interesting. It states that an increase in the market size does not translate in a larger number of actual applicants but only in higher applicants' ability. This is possible because competition discourages low ability employees from applying and thus, in equilibrium, only the better workers, who have more chances to be hired, are willing to bear the cost of sending an application. Therefore, when the market equilibrium implies separation, the self selection mechanism of applicants competing for a position is fully preserved. The following pictures show how the cost regions for which different equilibria arise change in k and N.

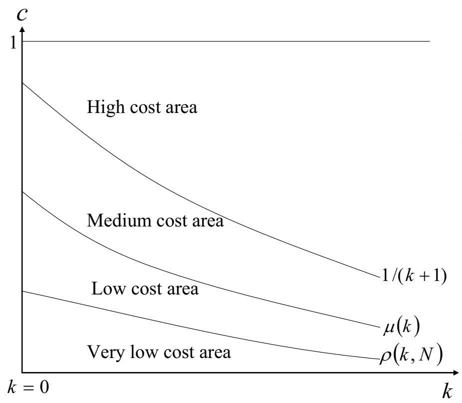


Figure 4(a)

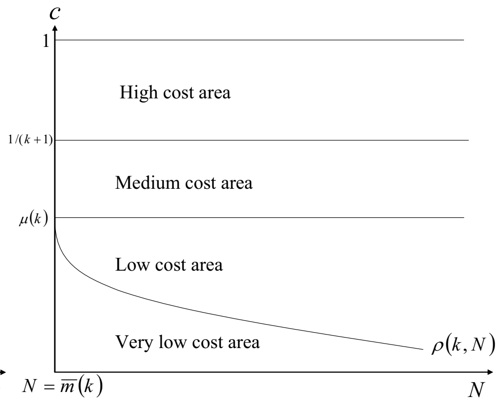


Figure 4(b)

On one hand, when the technology improves all cost regions, except the high-cost one, shrinks. Thus, for any sending cost c there exists a technology level that allows the Receiver to screen the best sender. On the other hand, a larger market size causes the very-low cost area to shrink and the low-cost to get larger. Thus, even for extremely low (but positive) sending cost, there exists a sufficiently large market size which ensures that the self selection mechanism plays a role. The reason is that, even though c lies initially in the very-low-cost area, which implies no separation, it will eventually lie in the low-cost area and, then, separation will occur.

The self selection mechanism has two very important implications on the functioning of the market. The first is that, from the Receiver point of view, a larger market is always better because the senders applying in equilibrium are of better quality. The second implication plays a role in determining the amount of information that the market generates, and it is addressed in the next section.

6 Market congestion and efficiency

In this section the analysis focuses on markets that suffer from information congestion, that is, markets where the cost of transmitting information is low relative to the information processing technology. Implicit here is the idea that the cost at which information can be transmitted is not a strategic variable, but is exogenously given by the environment. If, on the contrary, the Receiver could use the cost as a screening device (e.g. the Receiver could determine the sending cost by imposing an application fee) it would put it high enough to ensure that only the best sender finds worthwhile applying.

So far I have assume that the Receiver can neglect applications. Clearly this assumption does not capture the reality of many monopsony markets[8]. However, the model I propose can be easily accommodated to represent a situation in which neglecting applications is not an option. This can be done by restricting the strategy space of the Receiver to be simply that of selecting one applications from those received. The resulting model is much simpler as it implies that the Receiver maximizes expected utility $u_{\alpha}^{r}(\hat{x}, n)$ only with respect to $\hat{x}$, and that, the probability that a sender assigns to the event of being selected depends only on the total number of actual senders, $\phi_{\alpha}(\theta, n)$. Therefore, in this section, I will define and compare the economic cost of market congestion for two different scenarios: one in which applications can be neglected, and one in which this is not possible.

The first question I address here is how the amount of information generated, n, depends on the market size, N.

Proposition 2.4. *Let be $N > \bar{m}(k)$. For any $k \in (0, \infty)$ and $c \in (0, \mu(k))$ the market is overloaded and:*

($P.2.4.a$) *if the Receiver can neglect applications, the excessive amount of information is*

$$n(N) = \begin{cases} N & \text{if } \bar{m}(k) < N \leq \frac{1}{c}(1 - k\mu(k)) \\ \frac{1}{c}(1 - k\mu(k)) & \text{if } \frac{1}{c}(1 - k\mu(k)) < N < \infty \end{cases} \tag{11}$$

($P.2.4.b$) *if the Receiver cannot neglect applications, the excessive amount of information is*

$$n(N) = \begin{cases} N & \text{if } \bar{m}(k) < N \leq \frac{1}{c}(1 - kc) \\ \frac{1}{c}(1 - kc) & \text{if } \frac{1}{c}(1 - kc) < N < \infty \end{cases} \tag{12}$$

First of all, the proposition states that, irrespective of the fact that the Receiver is allowed or not allowed to neglect applications, the information generated is not strictly increasing in the market size. On the contrary, it is bounded, and its maximum level depends on c and k. It is the self selection mechanism that is responsible for this result. A positive cost on the supply side identifies a hard bound on the amount of information produced by the market: for N sufficiently large, separation occurs, consequently the maximum number of applicants in equilibrium is constant and depends only on c and k. Notice also that the technology level identifies a hard bound (that is $\bar{m}(k)$) on the maximum amount of information that the market is

[8]For instance, there are many situations in which the agent in charge of processing applications is legally constraint to pay attention to all applicants. For example scientific journals read all the submissions they receive.

willing to absorb. When $c < \mu(k)$, the former bound is larger than the latter, thus, the market is structurally subject to information congestion, the extent of which, is also bounded. Therefore the maximum magnitude of information overload, is $\frac{1}{c}(1 - k\mu(k)) - \bar{m}(k)$ if the receiver can neglect applications, and $\frac{1}{c}(1 - kc) - \bar{m}(k)$ if, on the contrary, applications cannot be neglected.

Since $c < \mu(k)$, from (11) and (12) it follows that the extent of market congestion is larger when the Receiver cannot neglect applications. Two effects play a role in shaping this result. First, if some applications are neglected in equilibrium, senders anticipate that there is a chance that their application might not be taken into account in the first place. Second, when applications are neglected, the signals' accuracy is larger compared to the case in which all excessive applications are processed, and this helps the Receiver to sort out relatively bad candidates. Both effects have a discouraging impact on senders and, therefore, a market in which neglecting applicants is an option induces less people to apply.

I now investigate what are the economic costs of information congestion. In an overloaded market in which some applications are neglected the inefficiency of information congestion arises only on the supply side. The Receiver, by ignoring some information, protects himself from the decreasing accuracy effect. However, since sending applications is costly, a market equilibrium with information overload implies too high sending costs for the society. Consider now an overloaded market in which all information generated must also be processed. Here, the side effects of information congestion are more severe and affect both the demand and the supply side. On one hand, the excessive number of applications processed decreases the utility of the Receiver[9]. On the other hand, the inefficiency is also present on the senders side as too many senders bear the cost of applying. Moreover, in a scenario in which applications cannot be neglected, the magnitude of the overload is larger compared to the situation in which ignoring applicants is an option. Therefore, here, the sending cost for the society are even higher. The next lemma is needed to support what follows.

Lemma 2.2. *Let be* $\Pi(k) = \pi(\bar{m}(k), k)$. *Then* $\Pi'(k) > 0$.

I now consider the total sending costs for the society. If applications can be neglected and the market is overloaded to its maximum extent, the total

[9] This is not true if the market size is very large. Indeed, the self-selection mechanism implies that, if $N \to \infty$, the threshold level α approaches 1, and, thus, irrespectively on the number of applications processed, the utility of the receiver also approaches 1.

sending cost for the society is

$$\begin{aligned} TC(c,k) &= c\left[\frac{1}{c}(1-k\mu(k))\right] \\ &= (1-k\mu(k)) = \left(1-\frac{k}{k+\bar{m}(k)}\right) = 1-\Pi(k) \qquad (13) \end{aligned}$$

If, on the contrary, applications cannot be neglected, the sending cost for the society is

$$TC(c,k) = c\left[\frac{1}{c}(1-kc)\right] = 1/c - k \qquad (14)$$

From (13) and (14) it follows that a marginal increase in the technology level always decreases the total sending cost. On one hand, a better technology implies that the Receiver can assess applicants ability more precisely, thus low ability applicants are discouraged from applying and the number of actual senders in equilibrium decreases. On the other hand, it also increases the maximum number of applications that the Receiver is willing to process. Consequently the extent of the overload decreases and, with it, also the total sending cost. Interestingly enough, in an equilibrium in which applications in excess are neglected, the total sending cost does not depend on the cost of sending one single application. The reason for this surprising result is that the elasticity of the number of actual senders with respect to the cost of sending one single application is always unitary: as the cost of sending one application increases the number of applicants in equilibrium decreases in the same proportion and therefore the total cost does not change. If, on the other hand, neglecting applications is not an option, the total sending costs are decreasing in the cost of sending one single application. Thus, in this second scenario, the elasticity of the number of actual senders with respect to the cost of sending one single application is negative and larger than one in absolute value.

7 Conclusion

Evidence of everyday life shows that there are many situations in which people receive too much information compared to what they are actually willing to process. In this chapter I addressed the information overload issue in the specific situation in which many applicants compete to obtain one position and the employer screens among applications in order to select the best applicant. With this set up information overload can be defined as an equilibrium outcome in which some applications are neglected by the economic agent in charge of screening applicants. It has been shown that a large market is not directly responsible for market congestion. On the contrary, more competition has a beneficial effect because the self-selection mechanism is preserved. The results I obtained show that information overload occurs when the application cost is low relative to the information processing technology level. This supports the intuitive idea that information overload is more likely to be present in environments in which the cost of transmitting information is low, while there are few resources to interpret information. I also noted that in some instances neglecting applications is not possible (e.g. because of a legal constraint) and, therefore, I also considerd a scenario in which all the information received must be processed. It turned out that the possibility of neglecting excessive information, to some extent, decreases the inefficiency of market congestion.

The general set up of the model is very natural and borrows standard concepts from the literature on economics of information. The crucial ingredient of the model is the assumption that when infinitely many applications are processed, then each application does not provide any information about the ability of the corresponding applicant. The more specific information processing technology that I used displays a clear trade-off between the quality and the quantity of information and is convenient for the possibility of obtaining closed form solutions with straightforward economic interpretations. More specifically, it allowed to interpret neglecting applications as a consequence of maximizing behavior where the marginal utility of a larger sample size (which is the higher chance of observing a very good application) is compared to its opportunity cost (which is the marginal decrease in the capability of ranking the applicants according to their true abilities). Yet, I believe that the main results are robust and hold true for more general information processing technologies.

8 Appendix to Chapter 2

Proof of lemma 2.1.

For notation simplicity, the subscript α of the distribution (3) will be omitted. I first show that $f(x) = I_{[\alpha,1]}(x)\frac{1}{1-\alpha}$. Since $F(\cdot \mid \theta, m)$ has a discrete jump at $x = \theta$, the difference $F(\theta \mid \theta, m) - \lim_{x\to\theta^-} F(x \mid \theta, m) = \pi(m)$ is the probability mass assigned to the event $x = \theta$, thus

$$\begin{aligned} f(x \mid \theta, m) &= I_{\{\theta\}}(x)\,\pi(m) + \frac{d}{dx}F(x \mid \theta, m) \\ &= \pi(m)\,I_{\{\theta\}}(x) + (1-\pi(m))\,I_{\{[\alpha,1]\setminus\{\theta\}\}}(x)\,\frac{1}{1-\alpha} \end{aligned}$$

Now

$$\begin{aligned} f(x \mid m) &= \int_{-\infty}^{+\infty} f(x \mid \theta, m)\,f(\theta)\,d\theta, \\ &= \frac{1}{1-\alpha}\int_{-\infty}^{+\infty} f(x \mid \theta, m)\,d\theta \\ &= \frac{1}{1-\alpha}\begin{pmatrix} \pi(m)\int_{-\infty}^{+\infty} I_{\{\theta\}}(x)\,d\theta + \\ +(1-\pi(m))\int_{-\infty}^{+\infty} I_{\{[\alpha,1]\setminus\{\theta\}\}}(x)\,\frac{1}{1-\alpha}d\theta \end{pmatrix} \\ &= \frac{\pi(m)}{1-\alpha}\int_{-\infty}^{+\infty} I_{\{\theta\}}(x)\,d\theta + \frac{1-\pi(m)}{1-\alpha}\int_{\alpha}^{1}\frac{1}{1-\alpha}d\theta \\ &= \frac{\pi(m)}{1-\alpha}\int_{-\infty}^{+\infty} I_{\{\theta\}}(x)\,d\theta + \frac{1-\pi(m)}{1-\alpha} \end{aligned}$$

Since $I_{\{\theta\}}(x) = 1$ when $\theta = x$, and since θ is uniformly distributed on $[\alpha, 1]$, $\int_{-\infty}^{+\infty} I_{\{\theta\}}(x)\,d\theta = \int_{\alpha}^{1}\frac{1}{1-\alpha}d\theta = 1$. It then follows that $f(x \mid m) = f(x) = I_{[\alpha,1]}(x)\frac{1}{1-\alpha}$. Now, $f(\theta \mid x, m) = \frac{f(\theta)f(x|\theta,m)}{f(x)}$ and, since $f(x) = f(\theta)$, we have $f(\theta \mid x, m) = f(x \mid \theta, m)$ and therefore

$$f(\theta \mid x, m) = \pi(m)\,I_{\{x\}}(\theta) + (1-\pi(m))\,I_{\{[\alpha,1]\setminus\{x\}\}}(\theta)\,\frac{1}{1-\alpha}$$

It then trivially follows that

$$\begin{aligned} E_\alpha[\theta \mid x, m] &= \pi(m)\,x + (1-\pi(m))\int_{-\infty}^{+\infty} I_{\{[\alpha,1]\setminus\{x\}\}}(\theta)\,\frac{\theta}{1-\alpha}d\theta \\ &= \pi(m)\,x + (1-\pi(m))\int_{\alpha}^{1}\frac{\theta}{1-\alpha}d\theta \\ &= \pi(m)\,x + (1-\pi(m))\left(\frac{\alpha+1}{2}\right) \end{aligned}$$

■

Proof of proposition 2.1.
I will use the following notation.

$$\begin{aligned} x_{(m)} &\equiv \max\{x_1, ..., x_m\} \\ \bar{x}_{(m)} &\equiv E_x[\max\{x_1, ..., x_m\}] \end{aligned}$$

First, the fact that $x^* = x_{(m)}$ is trivial given that $E_\alpha[\theta \mid x, m]$ is strictly increasing in x. Then, the reduced form payoff in stage 2 is

$$\begin{aligned} E_x\left[u^r\left(x_{(m)}, m\right)\right] &= \pi(m) E_x[\max\{x_1, ..., x_m\}] + (1 - \pi(m))\left(\frac{\alpha+1}{2}\right) \\ &= \pi(m) \bar{x}_{(m)} + (1 - \pi(m))\left(\frac{\alpha+1}{2}\right) \end{aligned}$$

I now show that $\bar{x}_{(m)} = \frac{\alpha+m}{m+1}$. First,

$$F_{x_{(m)}}(z) = \Pr(x_1 \le z, .., x_m \le z) = \Pi_{i=1}^m [F_{x_i}(z)]$$

Then, since signal outcomes are i.i.d. with $f(x) = I_{[\alpha,1]}(x)\frac{1}{1-\alpha}$, it follows that $F_{x_{(m)}}(z) = \left(\frac{z-\alpha}{1-a}\right)^m$ and $f_{x_{(m)}}(z) = \frac{m}{1-\alpha}\left(\frac{z-\alpha}{1-a}\right)^{m-1}$. Thus

$$\begin{aligned} \bar{x}_{(m)} &= \int_\alpha^1 f_{x_{(m)}}(z) z dz = \int_\alpha^1 \frac{m}{1-\alpha}\left(\frac{z-\alpha}{1-a}\right)^{m-1} z dz \\ &= \frac{\alpha+m}{m+1} \end{aligned}$$

Then, the reduced form pay-off is

$$\begin{aligned} E_x\left[u^r\left(x_{(m)}, m\right)\right] &= E_\alpha\left[\theta \mid \bar{x}_{(m)}, m\right] \\ &= \pi(m)\left(\frac{\alpha+m}{m+1}\right) + (1-\pi(m))\left(\frac{\alpha+1}{2}\right) \\ &= \left(\frac{\alpha+1}{2}\right) + \pi(m)\left(\frac{\alpha+m}{m+1} - \frac{\alpha+1}{2}\right) \\ &= \left(\frac{\alpha+1}{2}\right) + \pi(m)\frac{(1-\alpha)(m-1)}{2(m+1)} \end{aligned}$$

Notice that $E_x\left[u^r\left(x_{(1)}, 1\right)\right] = \lim_{m\to\infty} E_x\left[u^r\left(x_{(m)}, m\right)\right] = (\alpha+1)/2$, thus $\bar{m} \in (1, \infty)$. Let be $\gamma(m) = \pi(m)\frac{(1-\alpha)(m-1)}{2(m+1)}$, then

$$\gamma'(m) = \left(\frac{1-\alpha}{2}\right)\left(\frac{\pi'(m)(m-1)}{(m+1)} + \frac{\pi(m)2}{(m+1)^2}\right)$$

which yields the f.o.c.

$$\frac{|\pi'(\bar{m})|}{\pi(\bar{m})} = \frac{2}{\bar{m}^2 - 1} \tag{15}$$

The LHS of (15) captures the marginal cost of a larger sample size, while the RHS captures its marginal benefit. To show that $\bar{m}$ is unique, notice that, around $\bar{m}$, condition (15) can be approximated as follows

$$\begin{aligned}\frac{|\pi'(\bar{m})|}{\pi(\bar{m})+\pi'(\bar{m})(m-\bar{m})} &= \frac{2}{\bar{m}^2-1}\\ \frac{|\pi'(\bar{m})|}{\pi(\bar{m})} &= \frac{2}{(\bar{m}^2-1)+2(m-\bar{m})}\end{aligned}$$

The ratio $|\pi'(m)|/\pi(m)$ decreases in m at the rate of $1/m$, therefore, for any $m>\bar{m}$, the RHS of (13) is always larger then the LHS of (13) and the uniqueness of $\bar{m}$ follows. ■

Proof of lemma 2.2.

For notation simplicity, the subscript α of the distribution (3) will be omitted. Let y be the signal of sender θ, and $\{x_1,..,x_{m-1}\}$ the set of signals of the others $m-1$ senders. Let be $x_{(m-1)} \equiv \max\{x_1,..,x_{m-1}\}$. Recall from proof lemma 3.1 that the signal of an arbitrary sender is uniformly distributed on $[\alpha,1]$, that is,

$$F_x(z)=\frac{z-\alpha}{1-\alpha}$$

In the proof I will make use of the distribution of signals conditional on ability,

$$F_x(z\mid\theta,m)=I_{[\alpha,1]}(x)(1-\pi(m))F_x(z)+I_{[\theta,1]}(x)\pi(m)+I_{(1,\infty)}(x)$$

and the distribution of the maximum between $m-1$ signals coming from arbitrary senders

$$F_{x_{(m-1)}}(z)=[F_x(z)]^{m-1}$$

Then,

$$\begin{aligned}\phi_\alpha(\theta\mid m) &\equiv \Pr\left(x_{(m-1)}\le x\mid m,\theta\ge\alpha\right)\\ &= 1-\int_{-\infty}^{+\infty}F_x(z\mid\theta,m)f_{x_{(m-1)}}(z)\,dz\\ &= 1-\left((1-\pi(m))\int_\alpha^1 F_x(z)f_{x_{(m-1)}}(z)\,dz+\pi(m)\int_\theta^1 f_{x_{(m-1)}}(z)\,dz\right)\\ &= 1-\frac{(1-\pi(m))(m-1)}{(1-\alpha)^m}\int_\alpha^1(z-\alpha)^{m-1}\,dz-\\ &\quad -\frac{\pi(m)(m-1)}{(1-\alpha)^{m-1}}\int_\theta^1(z-\alpha)^{m-2}\,dz\\ &= 1-\left((1-\pi(m))\left(\frac{m-1}{m}\right)+\pi(m)\left(1-\left(\frac{\theta-\alpha}{1-\alpha}\right)^{m-1}\right)\right)\\ &= \pi(m)\left(\frac{z-\alpha}{1-\alpha}\right)^{m-1}+(1-\pi(m))\frac{1}{m}\end{aligned}$$

■

Proof of proposition 2.2.

Let α be the threshold level of a PSE, let $n = (1-\alpha) N$ be the number of senders applying in equilibrium. let $\Psi(\alpha, N, k)$ be the probability that the sender located at α assigns to the event of being selected in equilibrium. Notice first that, in equilibrium, it must be $n \geq 1$. If $n < 1$, then the mass of applying senders is smaller than the mass of receivers, and some receivers do not receive any application. Therefore provided $n < 1$, some senders have an incentive to apply because they know they will be selected for sure. Recalling that $N > \bar{m}(k)$ and noticing that, by definition, $\Psi(\alpha, N, k) \equiv \Phi_\alpha(\alpha \mid n, m^*(n))$, it follows that

$$\begin{aligned} \Psi(\alpha, N, k) &= \begin{cases} 1 & \text{if } \ 1 = n \\ \phi_\alpha(\alpha, n) & \text{if } \ 1 < n \leq \bar{m}(k) \\ \frac{m(k)}{n}\phi_\alpha(\theta, \bar{m}(k)) & \text{if } \ \bar{m}(k) < n < N \end{cases} \\ &= \begin{cases} 1 & \text{if } \ \alpha = \frac{N-1}{N} \\ \frac{1}{k+(1-\alpha)N} & \text{if } \ \frac{N-\bar{m}(k)}{N} \leq \alpha < \frac{N-1}{N} \\ \frac{\bar{m}(k)}{(1-\alpha)N}\frac{1}{k+\bar{m}(k)} & \text{if } \ 0 \leq \alpha < \frac{N-\bar{m}(k)}{\bar{m}(k)} \end{cases} \end{aligned} \tag{16}$$

Since $\mu(k) \equiv 1/(k + \bar{m}(k))$ and $\rho(k, N) \equiv (m(k)/N)\mu(k)$, the functional form (16) proves the proposition. ■

Proof of corollary 2.1.

The fact that a market is overloaded when $c < \mu(k)$ is trivial and follows directly from (16). I then show that $\mu'(k) < 0$.

$$\mu(k) = 1/(k + \bar{m}(k))$$

$\bar{m}(k)$ is given by the f.o.c. (15)

$$\begin{aligned} \frac{|\pi'(\bar{m}, k)|}{\pi(\bar{m}, k)} &= \frac{2}{\bar{m}^2 - 1} \\ \frac{1}{k + \bar{m}} &= \frac{2}{\bar{m}^2 - 1} \\ &\implies \bar{m}(k) = 1 + (2(k+1))^{1/2} \end{aligned} \tag{17}$$

therefore, $\mu'(k)$ and $\lim_{k\to\infty}\mu(k) = 0$. ■

Proof of proposition 2.3.

Two cases are considered.

Case 1: $n \leq \bar{m}(k)$. The equilibrium condition is

$$\frac{1}{k + (1-\alpha) N} - c = 0 \tag{18}$$

and implicit differentiations on (18) yield $\partial\alpha/\partial k = 1/N > 0$ and $\partial\alpha/\partial N = (1-\alpha)/N > 0$. It then follows that

$$\begin{aligned}
\partial n/\partial k &= -(\partial\alpha/\partial k)\, N = -1 < 0 \\
\partial n/\partial N &= (1-\alpha) - N(\partial\alpha/\partial N) = (1-\alpha) - N\left(\frac{1-\alpha}{N}\right) = 0
\end{aligned}$$

Case 2: $n > \bar{m}(k)$. The equilibrium condition is

$$\begin{aligned}
\frac{\bar{m}(k)}{(1-\alpha)\,N\,(k+\bar{m}(k))} - c &= 0 \\
\frac{\bar{m}(k)\,\mu(k)}{(1-\alpha)\,N} - c &= 0 \qquad (19)
\end{aligned}$$

Recalling (17), and applying implicit differentiations on (19), yields

$$\begin{aligned}
\partial\alpha/\partial k &= -\frac{(1-\alpha)\,N}{N\bar{m}(k)\,\mu(k)}\left[\bar{m}'(k)\,\mu(k) + \bar{m}(k)\,\mu'(k)\right] \\
&= \frac{(1-\alpha)\,N}{N\bar{m}(k)\,\mu(k)}\left[\frac{2+k+(2(k+1))^{1/2}}{(2(k+1))^{1/2}\left(1+k+(2(k+1))^{1/2}\right)^2}\right] > 0 \\
\partial\alpha/\partial N &= \frac{1-\alpha}{N} > 0
\end{aligned}$$

It then follows

$$\begin{aligned}
\partial n/\partial k &= -(\partial\alpha/\partial k)\, N < 0 \\
\partial n/\partial N &= (1-\alpha) - N(\partial\alpha/\partial N) = (1-\alpha) - N\left(\frac{1-\alpha}{N}\right) = 0
\end{aligned}$$

■

Proof of proposition 2.4.

If the receiver can neglect applications, then $\Psi(\alpha, N, k)$ is given by (16). Since $c < \mu(k)$, and $\bar{m}(k) > N$, if

$$\begin{aligned}
0 &< c < \rho(k,N) \equiv \frac{\bar{m}(k)}{N\,(k+\bar{m}(k))} \\
N &< \frac{\bar{m}(k)}{c\,(k+\bar{m}(k))} = \frac{1}{c}\left(1 - \frac{k}{k+m(k)}\right) = \frac{1}{c}\,(1-k\mu(k))
\end{aligned}$$

the PE arises and $n = N$. If,

$$\begin{aligned}
c &\geq \rho(k,N) \equiv \frac{\bar{m}(k)}{N\,(k+\bar{m}(k))} > 0 \\
N &\geq \frac{\bar{m}(k)}{c\,(k+\bar{m}(k))} = \frac{1}{c}\left(1 - \frac{k}{k+m(k)}\right) = \frac{1}{c}\,(1-k\mu(k))
\end{aligned}$$

a PSE arises and, from $(P.5.2.b)$, n is constant for any marginal increases in N. Therefore the (bounded) amount of information is given by the number of senders applying in the PSE for which the threshold level is $\alpha = 0$, that is, n such that

$$\begin{aligned} c &= \frac{\bar{m}(k)}{n(k+\bar{m}(k))} \\ n &= \frac{\bar{m}(k)}{c(k+\bar{m}(k))} = \frac{1}{c}(1-k\mu(k)) \end{aligned}$$

If, on the other hand, the Receiver cannot neglect applications, then $\Psi(\alpha, N, k)$ is simply

$$\begin{aligned} \Psi(\alpha, N, k) &= \begin{cases} 1 & \text{if } \ 1 = n \\ \phi_\alpha(\alpha, n) & \text{if } \ 1 < n \leq N \end{cases} \\ &= \begin{cases} 1 & \text{if } \ \alpha = \frac{N-1}{N} \\ \frac{1}{k+(1-\alpha)N} & \text{if } \ 0 \leq \alpha < \frac{N-1}{N} \end{cases} \end{aligned}$$

Thus, if

$$\begin{aligned} 0 &< c < \frac{1}{k+N} \\ N &< \frac{1}{c}(1-kc) \end{aligned}$$

the PE arises and $n = N$. If, on the contrary,

$$\begin{aligned} \frac{1}{k+N} &\leq c \\ N &\geq \frac{1}{c}(1-kc) \end{aligned}$$

a PSE arises and n is constant for any marginal increases in N. Therefore the (bounded) amount of information is given by the number of senders applying in the PSE for which the threshold level is $\alpha = 0$, that is, n such that

$$\begin{aligned} c &= \frac{1}{k+n} \\ n &= \frac{1}{c}(1-kc) \end{aligned}$$

■

Proof of Lemma 2.2.

$$\begin{aligned} \Pi(k) &\equiv \frac{k}{k+\bar{m}(k)} \\ &= \frac{k}{k+1+(2(k+1))^{1/2}} \end{aligned}$$

Thus,

$$\Pi'(k) = \frac{2 + k + (2(k+1))^{1/2}}{(2(k+1))^{1/2}\left(1 + k + (2(k+1))^{1/2}\right)^2} > 0$$

■

3. The effect of lack of communication and limited attention on organizational strategies[1]

1 Introduction

Since the analysis of this chapter is focused on collective decision-making, by imperfect information processing I mean the processing limitation of the decision-making network (i.e., lack of communication). Yet, in some sections of the chapter I also investigate the effect of individual processing limitations (i.e., lack of attention) on the collective decision.

Let us consider the following example. Suppose a car manufacturer has to choose one from many potential prototype projects in order to launch a new car model. The selection criteria must pay attention to both technical (performance) and esthetical (design) considerations. Since the company policy gives high priority to safety and quality standards, the selection procedure is organized as follows. First, a group of engineers from R&D department tests the technical performance of each available project and, based on the information gathered, selects, let us say, n best projects to be further investigated. Then, a group of designers from the marketing department compares those projects and, depending on how "good" they look, selects the model that will be eventually launched in the market. They are two types of complications that make the decision-making in this example very difficult. The first comes from the consideration that companies have limited budgets and, consequently, the resources each department can allocate to screen projects are limited. In our example, it is reasonable to assume that, if the marketing department is left with too many projects, it will not be able to rank them very accurately. Therefore, the actual number of projects pre-selected by the R&D department may well affect the overall accuracy of the selection procedure. The second complication arises from the fact that communication is always imperfect. This is due to inevitable information contamination, or to a high degree of specialization, which makes it hard for people with different backgrounds to understand each other. In the example, it is a necessity to combine both performance and design characteristics into a single measure that requires information sharing between the two departments. Hence, the overall quality of selection crucially depends on information that is actually communicated between R&D and marketing departments.

This example clearly shows that internal information-processing limitations may prevent an otherwise perfect decision-maker from selecting the

[1] This chapter is based on a joint work with Vladimir Karamychev.

best feasible alternative with certainty. Organizations of individuals have constraints in both acquiring and communicating information. Quantitatively, more information means more alternatives and this requires more resources to compare all alternatives. Qualitatively, more complex information means that different expertises are needed to evaluate each alternative. In accordance with this dichotomy, I distinguish here two information-processing limitations in. The first one manifests itself in a sample size accuracy trade-off: the more alternatives are simultaneously processed, the smaller is the accuracy with which each alternative is evaluated. I will refer to this limitation as limited attention. The second limitation plays a role on the information transmission level due to imperfect communication between members of the same organization. Since a decision-maker can be thought as an information–processing network, this limitation can be interpreted as imperfect transmission from one node of the decision-maker's internal structure to another. I will refer to this limitation as imperfect communication.

The aim of this chapter is to incorporate both these information-processing limitations in a model of rational individuals. More precisely, I consider a two-stage selection procedure with two selectors, who evaluate an exogenous number of alternatives in order to select the best one. Since selectors can be seen as two team members of the same organization (in the sense of Marschak and Radner (1972)), the selection problem can be naturally modeled as a two-stage game where two players have the same preferences over the outcomes of the game.

The game is as follows. First, nature assigns the types to each alternative. For the sake of simplicity, alternatives are assumed to come into two types: high or low. In stage 1, each alternative generates an imprecise binary (high or low) signal about its quality. Having observed signals' outcomes, selector 1 selects a sub-sample of alternatives to be passed to the next stage. In stage 2, each pre-selected alternative generates another imprecise binary signal, and selector 2 selects one out of them. The payoff of each selector is the probability that a high quality alternative is selected in stage 2.

In terms of the model, the example is formalized as follows. Each project can be of either high or low quality. Both R&D and marketing departments are given criteria that they apply to projects. First, the R&D department screens all the projects. It observes a high signal if a project meets the requirement and a low signal if otherwise. Based on these observations, the R&D department selects a sub-sample of projects and passes it to the marketing department. Similarly, the latter applies its own criterion to each alternative in the sub-sample and observes a high signal if a project meets the requirement.

In general, each type of alternatives may generate either a high or a low signal in every stage. However, I also consider two special cases of the model, where results can be generalized to an arbitrary number of stages. In the first case, I assume that a low-type alternative can never generate a high signal.

This corresponds to a situation in which the screening requirements are set so high that no low quality alternative can ever meet them. I call this case high-standard filtering selection as only high types can pass screening filters by generating high signals. The opposite case, is one where I assume that a high type alternative never generates a low signal. This corresponds to setting the screening requirements so low that every high quality alternative meets them for sure. I call this case low-standard filtering selection as only low types may fail to pass screening filters by generating low signals.

In my analysis I consider four cases. The first one is a benchmark case where the two information-processing limitations are absent: attention is unlimited and communication is perfect. This situation is captured by assuming that all information obtained at stage 1 (the signal outcomes that each alternative generated in stage 1) is available at stage 2, and that the accuracy of signals in stage 2 is constant. I refer to this case as perfect information environment case. This benchmark scenario is reminiscent of a perfect statistical environment where handling large samples is not costly and all the information gathered from sequential experiments could be used.

In order to analyze the effect of each of the information-processing limitations on the behavior of rational agents, I then depart from the benchmark scenario by analyzing three other cases. In the first case, attention is unlimited, but the assumption of perfect communication is relaxed by assuming that no information gathered in stage 1 is available in stage 2. I refer to this case as no-communication case. In the second scenario, on the contrary, communication is kept perfect, but attention is assumed to be limited. I assume there, that the signals' accuracy in stage 2 is decreasing in the number of alternatives to be evaluated, i.e., in the sample size in stage 2. I refer to this case as limited attention case. Lastly, combining both types of imperfections I analyze a fourth scenario where there is no communication and attention limited.

It turns out that in any scenario, regardless of the underlying informational assumptions, multiple Nash equilibria exist. That is why I treat the problem of selecting the best alternative from a game-theoretic perspective rather than from a purely statistical point of view. As the initial types of alternatives chosen by nature is not observable, the game has only one subgame – the game itself – , and the subgame perfection does not reduce a set of equilibria. On the other hand, the notion of weak perfect Bayesian equilibrium requires a specification of players' beliefs in all stages and all possible signals' realizations that makes this equilibrium notion too cumbersome. Instead, I use the normal form trembling-hand perfect Bayesian notion (PBE thereafter) and it turns out that such equilibrium always exists and is unique for all generic values of the model's primitives.

The results are as follows. The PBE in the perfect information environment case is such that independently of the signal outcomes in stage 1, all alternatives are passed to stage 2. Selector 2 makes use of both signals of each alternative in order to select the best one. This equilibrium captures a well-known concept in statistics: calculate likelihoods of all possible alternatives using all available information, and then select an alternative with the largest likelihood value.

When I depart from the benchmark perfect information scenario, the paradigm "more information is better" does not hold true any longer. Irrespective of the sources of imperfection, in some cases selector 1 is better off by neglecting a part of potentially valuable information. The origin of this phenomenon depends on information-processing limitations.

Assuming no communication, not surprisingly, reduces incentives of selector 1 to select both high and low signals into a single pool as such mixing makes it impossible for selector 2 to distinguish between them later on. However, mixing does occur in equilibrium, although only a subset of the low signals is selected. The PBE for the no communication case has the following properties: all high signals in stage 1 are passed to stage 2, while some low signals may not be selected. Thus, a lack of communication puts a bound on the number of relatively bad alternatives to be passed to the next selection round. This upper-bound of the number of low signals is decreasing in the number of high signals observed, in the prior share of high type alternatives in the population, and in the screening accuracy in stage 1.

When there is limited attention, it is the decrease in accuracy in stage 2 that prevents selector 1 from taking too many alternatives. The PBE in this case has the following properties: either all high signals are selected but some low signals are neglected, or some high signals are neglected and no low signals are selected. Thus, limited attention puts a bound on the total number of alternatives, either good or bad, to be passed to the next selection round. Contrary to the no communication case, in the limited attention case the screening accuracy in stage 1 has an ambiguous impact on the upper-bound of the number of low signals while both the number of high signals observed and the prior share of high type alternatives in the population affect it negatively and monotonically.

In the fourth scenario, when both sources of informational imperfections are present, the upper-bounds of the numbers of high and low signals are even lower as both imperfections work hand-in-hand in the same direction.

Finally, in multi-stage filtering selection I obtain the following results. In case of high-standard filtering selection neglecting information is never optimal provided the number of alternatives is sufficiently large. The initial set of alternatives passes through screening filters on every stage until at least one high signal is generated and one of the corresponding alternatives is selected. If, on the other hand, the standards are low, then only high signals are passed to the next stage. In equilibrium, very few alternatives will be selected in each stage. More precisely, only two high signals are passed to the next stage provided the alternatives have passed a sufficiently large number of filtering stages.

The rest of the chapter is organized as follows. Section 2 reviews the existing related literature. Section 3 states the model. Section 4 analyzes four different informational scenarios. Section 5 analyzes two special filtering selection cases and section 5 concludes. Section 6 is the appendix and contains all the proofs.

2 Related literature

Collective decision-making is usually modeled by viewing an organization as an information-processing network where each node corresponds to an agent belonging to the organization. To this respect this chapter stands in the tradition of Marschak and Radner (1972) as selectors are modeled as team members and therefore agency problems are ignored. The current chapter shares some feature with at least two different approaches that economic literature has taken in modeling economic organizations. First, similarly to Radner (1993)[2], the analysis focuses on a hierarchical structure whose task is to select the best out of several available alternatives. In Radner (1993) imperfect information processing refers to the fact that, due to limited attention, an agent can process (perfectly) only a limited number of alternatives within a time unit. Consequently, the gaol is to characterize the efficient hierarchical structure, that is, the hierarchy which selects the best alternative in the shortest possible time and by using the smallest possible number of agents. Differently I model limited attention by assuming trade off between sample size and screening accuracy, and the goal of my analysis is to characterize the effect that such information-processing limitation has on the optimal selection strategy. Second, I share with Sah and Stiglitz (1985, 1986), Koh (1992b) and Visser (2000) the idea that humans are fallible screeners (i.e. they make decisions based on imprecise information) and that they cannot communicate the information they posses. Differently from the above mentioned authors, the task of the organization is not that distinguishing good alternatives from bad alternatives, but is that of selecting the best alternative from a given set.

Finally, in this chapter I share with Moscarini and Smith (2002) the interpretation of the amount of information as the number of signal outcomes drawn.

3 The model

There are two selectors and a population of N alternatives, which come into two types: high-type and low-type, denoted as θ_H and θ_H respectively. The expected share of θ_H within the population is denoted by α. The

[2] Other authors have taken the same approach of Radner. See, for instance, Meagher and Van Zandt (1998), Radner and Van Zandt (1992, 2001), and Van Zandt (1998, 1999).

game lasts two stages. In stage 1 each alternative generates a binary signal $x_1 \in \left\{x_1^H, x_2^L\right\}$, i.e., either a high signal or a low signal, which is correlated with its true type with the following revealing probabilities:

$$\Pr\left(x_1^H \mid \theta_H\right) = q_1^H \in (0,1), \quad \Pr\left(x_1^L \mid \theta_L\right) = q_1^L \in (0,1)$$

Having observed a signal composition (H_1, L_1), i.e. number H_1 of high signals and a number L_1 of low signals, selector 1 selects a sub-sample of alternatives (h_1, l_1) which consists of $h_1 \in [0, H_1]$ high signals and $l_1 \in [0, L_1]$ low signals.

In stage 2, each pre-selected alternative again generates a binary signal $x_2 \in \left\{x_1^H, x_2^L\right\}$ in accordance with the revealing probabilities $q_1^H \in (0,1)$ and $q_1^L \in (0,1)$. When communication is perfect, selector 2 makes his choice based on two signals, x_1 and x_2, observed in both stages. The pair (x_1, x_2) determines the overall likelihood value for each alternative. When there is no communication, on the contrary, the signaling history is not available, and selector 2 makes his choice based on signals x_2 only. Having observed realizations of signals in stage 2, selector 2 selects one alternative, which becomes the outcome of the selection procedure. I assume that all signals in stages 1 and 2 are statistically independent:

Assumption 1. $\Pr(x_1, x_2 \mid \theta_i) = \Pr(x_1 \mid \theta_i) \Pr(x_2 \mid \theta_i), \quad i = H, L.$

For binary types and signals, the labeling of the types can always be done in such a way, that the monotone likelihood ratio property holds: high signal x_t^H gives more chances of being generated by a high type alternative. Formally, I assume that it is indeed the case.

Assumption 2.

$$\begin{aligned} \gamma_t^H &\equiv \Pr\left(\theta_H \mid x_t^H\right) = \frac{\alpha_t q_t^H}{\alpha_t q_t^H + (1-\alpha_t)\left(1-q_t^L\right)} > \\ &> \frac{\alpha_t\left(1-q_t^H\right)}{\alpha_t\left(1-q_t^H\right) + (1-\alpha_t) q_t^L} = \Pr\left(\theta_H \mid x_t^L\right) \equiv \gamma_t^L \end{aligned}$$

The information acquisition technology in the model is represented by the revealing probabilities q_t^H and q_t^L. The screening accuracies in stage 1, q_1^H and q_1^L are exogenously given. In contrast, the screening accuracies in stage 2, q_2^H and q_2^L, are determined by the sample size $N_2 = h_1 + l_1$: If attention is unlimited q_2^H and q_2^L are constant functions while for the case of limited attention they are strictly decreasing in N_2.Since, when attention is limited, the revealing probability functions $q_2(n)$ are strictly decreasing and bounded, they can be written as $q_2(n) = \underline{q} + f(1/n)$ where $\underline{q} \geq 0$ and $f(x)$ is strictly increasing function such that $f(x) = 0$. Function f can be treated as a production function of information acquisition technology and its argument $1/n$ represents the amount of resources allocated for each alternative. We assume that $q_2(n)$ is a well-behaved function for all large enough values of its argument, i.e., $f(x)$is a well-behaved function at zero.

Assumption 3. $\underline{q} > 0$ and $f(x)$can be written as $f(x) = x^{\lambda}g(x)$ for some$\lambda > 0$ and an arbitrary function $g(x)$, which is differentiable at $x = 0$ and satisfies g(0) > 0.

The pay-off of each player is the probability that a high type alternative is eventually selected. An equilibrium strategy for selector 1 is the optimal sample composition $(h_1^*, l_1^*) = (h_1^*(H_1, L_1), l_1^*(H_1, L_1))$ for all possible signal compositions (H_1, L_1). An equilibrium strategy for selector 2 is to select an alternative in accordance with his preference relation by comparing likelihoods of each alternative.

4 Analysis

Due to the finiteness of the strategy space of the game, a Bayesian equilibrium always exists, possibly in mixed strategies. Moreover, a pure strategy Bayesian equilibrium always exists as the players are team members and their pay-off functions coincide. In what follows, I will consider Bayesian equilibria in pure strategies only.

The model, regardless of the underlying informational assumptions, has always multiple Nash equilibria. In order to see why this is the case, consider the following strategy profile: selector 1 passes only one signal, preferably high, to stage 2; selector 2 selects one alternative with the lowest likelihood of being θ_H type. Given the strategy of player 2, player 1 wants to end the selection procedure in stage 1 by making the team's pay-off independent on the signal realization in stage 2. Thus, he optimally selects only one alternative, and it must be one that generated a high signal , x_1^H, if there is one. Player 2, in turn, gets the same pay-off irrespective of his strategy, thus there is no profitable deviation for him.

It is clearly seen that the equilibrium we just described is based on playing weakly dominated strategies. That is why we impose an additional refinement, namely that no weakly dominated strategies are a part of an equilibrium. The set of strategies that are not weakly dominated can be characterized as follows: selector 1 (2) selects a number of signals (one signal) with the largest likelihood(s) of being θ_H type.

We begin with the benchmark case, where the informational environment, apart from inaccurate screening, is perfect.

4.1 Perfect communication and unlimited attention

Without informational imperfections, there are no costs of passing large samples to stage 2. In addition, selector 2 observes all signals from stage 1. Hence, he can rank all the previously selected alternatives in accordance with its preference relation. Thus, selector 2 has a unique weakly undominated strategy. Selector 1, in turn, selects all potentially valuable alternatives in stage1. In addition, in a perfect equilibrium, he must not select any other alternatives. This is the content of Proposition 3.1.

Proposition 3.1 *In perfect information environment case, the game has a unique PBE such that:*

a) $h_1^*(H_1, L_1) = H_1$, *i.e., player 1 selects all high signals.*

b) $l_1^*(H_1, L_1) = \begin{cases} L_1 & \text{if } \frac{q_2^H}{1-q_2^H}\frac{q_2^L}{1-q_2^L} > \frac{q_1^H}{1-q_1^H}\frac{q_1^L}{1-q_1^L} \\ 0 & \text{if } \frac{q_2^H}{1-q_2^H}\frac{q_2^L}{1-q_2^L} \leq \frac{q_1^H}{1-q_1^H}\frac{q_1^L}{1-q_1^L} \end{cases}$

i.e., player 1 selects all low signals if the screening accuracy in stage 2 is higher than in stage 1 and none of them otherwise.

c) Player 2 selects an alternative in accordance with the following preference relation

$\left(x_1^H, x_2^H\right) \succ \left(x_1^L, x_2^H\right) \succ \left(x_1^H, x_2^L\right) \succ \left(x_1^L, x_2^L\right)$ *if* $\frac{q_2^H}{1-q_2^H}\frac{q_2^L}{1-q_2^L} > \frac{q_1^H}{1-q_1^H}\frac{q_1^L}{1-q_1^L}$

$\left(x_1^H, x_2^H\right) \succ \left(x_1^H, x_2^L\right) \succ \left(x_1^L, x_2^H\right) \succ \left(x_1^L, x_2^L\right)$ *if* $\frac{q_2^H}{1-q_2^H}\frac{q_2^L}{1-q_2^L} \leq \frac{q_1^H}{1-q_1^H}\frac{q_1^L}{1-q_1^L}$

Proposition 3.1 is proven as a sub-case of Proposition 3.3 in the appendix. It can be easily generalized to an arbitrary number of stages. Proposition 3.1 states one of the most general concepts in statistics that prescribes a calculation of full likelihoods for all available alternatives, and then selection of the maximum value. When the screening accuracy in stage 1 is higher than in stage 2, i.e., when $\Pr\left(\theta_H \mid x_1^H, x_2^L\right) > \Pr\left(\theta_H \mid x_1^L, x_2^H\right)$, and, therefore, $\left(x_1^H, x_2^L\right) \succ \left(x_1^L, x_2^H\right)$, only high signals are selected. If, on the contrary, stage 2 signaling is more accurate, i.e., $\Pr\left(\theta_H \mid x_1^H, x_2^L\right) \leq \Pr\left(\theta_H \mid x_1^L, x_2^H\right)$, selector 1 selects all the alternatives. It is important to note that the weak dominance criterion does not yield a unique equilibrium. The equilibrium described in Proposition 3.1 is the only perfect equilibrium out of all Bayesian equilibria in weakly undominated strategies[3].

Having established the result in the benchmark case, I will see now how informational imperfections affect the resulting equilibrium strategies. The first imperfection I introduce is lack of communication.

[3] If the accuracy in stage 1 is higher than in stage 2, then selecting all alternatives in stage 1 is also an equilibrium.

4.2 Lack of communication

If no information from stage 1 is available for selector 2, he still has a unique strategy but now based only on the following preference relation over the signals from stage 2:

$$\left(x_2^H\right) \succ \left(x_2^L\right) \tag{1}$$

Selector 1 now faces the following task. Observing a signal composition (H_1, L_1) he has to choose a sample composition (h_1, l_1) to be passed to stage 2 in order to maximize the team's pay-off. An equilibrium sample composition is denoted by (h_1^*, l_1^*) Proposition 3.2 shows how the realization of signals in stage 1 affects the optimal sample composition.

Proposition 3.2 *In the no communication case there exists a generically unique PBE such that:*

a) $h_1^*(H_1, L_1) = H_1$*, i.e., player 1 selects all high signals.*
b) for all $H_1 \geq 1$ *there exist an upper-bound* $\bar{L}_1(H_1) \in [0, \infty)$ *and a lower-bound* $\underline{L}_1(H_1) \in [0, \bar{L}_1(H_1)]$ *such that:*

$$l_1^*(H_1, L_1) = \begin{cases} \min\left\{L_1, \bar{L}_1(H_1)\right\} & \text{if } L_1 \geq \underline{L}_1(H_1) \\ 0 & \text{if if } L_1 < \underline{L}_1(H_1) \end{cases}$$

i.e., if the number of low signals does not exceed $\underline{L}_1(H_1)$ *none of them are selected; otherwise all of them up to* $\bar{L}_1(H_1)$ *are selected in stage 1.*
c) $\bar{L}_1(H_1)$ *does not increase and strictly decreases whenever* $\bar{L}_1(H_1) > 0$*.*
d) There exists a number $\tilde{H}_1 \leq L_1(1) + 1$ *such that* $l_1^*(h, L_1) = \underline{L}_1(h) = \bar{L}_1(h) = 0$ *for all* $h \geq \tilde{H}_1$*, i.e., if there are sufficiently many high signals, none of low signals are selected in stage 1.*
e) $l_1^*(h, L_1) = \bar{L}_1(H_1) = \underline{L}_1(H_1) = 0$ *for all* $h \geq 1$ *if* $\frac{q_2^H}{1-q_2^H}\frac{q_2^L}{1-q_2^L} \leq \frac{q_1^H}{1-q_1^H}\frac{q_1^L}{1-q_1^L}$*, i.e., if the screening accuracy in stage 1 is higher than in stage 2 none of low signals are selected in stage 1.*
f) $l_1^*(0, L_1) = L_1$*, i.e., if there are no high signals available, all low signals are selected. Formally,* $\bar{L}_1(0) = +\infty$*,* $\underline{L}_1(0) = 0$
g) The total sample size in stage 1, $h_1^* + l_1^*$ *is not a strictly monotone function of* H_1*: it weakly decreases when* $l_1^* > 0$ *and strictly increases when* $l_1^* = 0$*.*
h) Player 2 selects an alternative in accordance with the preference relation (1).

The statements of Proposition 3.2 can be understood as follows. First, PBE is generically unique as the pay-off function of player 1 turns out to be a regular analytical function of the model's primitives and, therefore, takes generically different values for all different values of its arguments, which number is finite. Then, selecting all high signals is always optimal due to Assumption 2, part (a). Selecting one extra low signal has two effects on the pay-off. The first effect, which is positive, is a sample size effect: selecting more signals in stage 1 increases the probability of observing at least one high signal in stage 2. The other effect, which is a mixing effect, is negative due to the no communication assumption. Mixing high and low signals in stage 1 in a single pool makes it impossible to distinguish between them later on in stage 2 and, therefore, decreases the probability of selecting the best alternative.

The sample size effect vanishes exponentially with the number of selected low signals while the mixing effect decreases reciprocally to that number. Hence, for a large number of low signals selected the latter dominates the former and there is an upper bound $\bar{L}_1$ such that $l_1^* \leq \bar{L}_1$. If only a few low signals are available, it might be possible that taking none of them is optimal even if $\bar{L}_1 > 0$ as the mixing effect is absent in this case. Thus, the existence of the lower bound $\underline{L}_1$ is established, part (b).

When the number of high signals goes up, the sample size effect vanishes faster then the mixing effect. Therefore, selector 1 has less incentives to select low signals. As a result, the upper-bound $\bar{L}_1$ strictly decreases w.r.t. H_1 until it becomes zero, and stays at zero afterwards, part (c) and (d). Selector 1 selects high and low signals only if the screening accuracy in stage 1 is lower than in stage 2. Otherwise none of low signal will be selected, part (e).

When there are no high signals available, only the sample size effect plays a role and, therefore, all low signals must be selected, part (f). Finally, the total sample size in stage 1 cannot be a strictly monotone function as for $H_1 = 0$ and $H_1 = N$ all signals will be selected, thus $h_1^* + l_1^* = N$ in these two cases.

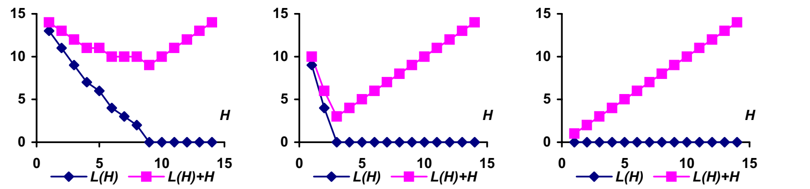
As we see, departing from the benchmark case by imposing no communication, not surprisingly, leads to inefficiency: in large heterogeneous samples a part of potentially valuable alternatives that generated low signals, will be discarded. However, it is important to note that it is not the amount of information (i.e., the number of alternatives) passed to the second stage but rather its diversity (i.e., signals' heterogeneity) that creates this phenomenon.

Only a part of low signals will be selected in equilibrium and, therefore, only a part of all available information will be used in the decision-making. Due to the non-monotone behavior of the total sample size, $N_2 = h_1^* + l_1^*$, the value N_2 does not fully reveal the sample composition (how many high and low signals have been selected in stage 1). Thus, in a multistage ($T > 2$) selection game, selector 2 faces a non-trivial task of updating sample composition beliefs, which makes the model practically intractable for $T > 2$.

For exposition purposes, we have numerically calculated the function in the following example.

Example 1 *Picture 1 shows the numerically calculated function* $\bar{L}_1(H_1)$ *for* $\alpha = 0.5$ *and three values of* $q_1^H = q_1^L$*: 0.51, 0.65 and 0.78. Its monotone property can be easily seen there, as well as the non-monotone behavior of the maximum sample size* $\bar{L}_1(H_1) + H_1$*. Picture 1 also shows how the accuracy in stage 1 affects the sample composition. If the signaling stage 1 is almost uninformative, picture (a), the sample size effect is very large and selector 1 aggressively mixes signals for a wide range of* H_1*. If, on the contrary, the accuracy in stage 1 is sufficiently high, picture (c), the mixing effect dominates and selector 1 always neglects low signals. Another feature of the equilibrium is that player 1 selects "very many" low signals only if the prior is low, the accuracy in stage 1 is low and the number of high signals available is also low.*

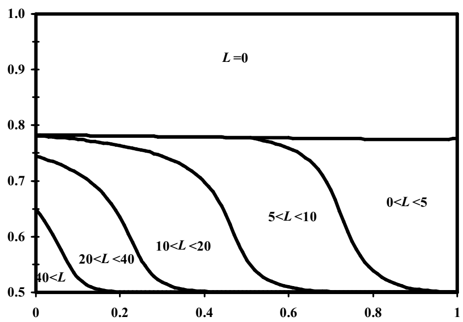
Picture 2 shows regions of the prior α *and the first stage accuracy* $q_1^H = q_1^L$ *where* $L = \bar{L}_1(1)$ *takes different values, for the case* $q_1^H = q_1^L = 0.9$*. One may note that both the prior and the first stage accuracy monotonically and negatively affect the upper-bound* $\bar{L}_1(1)$*. This monotone dependence of* $\bar{L}_1(H_1)$ *is confirmed by numerous numerical calculations, yet the analytical proof remains to be found.*



a) $q_1 = 0.51$ b) $q_1 = 0.65$ c) $q_1 = 0.78$

Picture 1

The upper-bound $\overline{L_1}$, denoted as $L(H)$, and the maximum sample size in stage 1, denoted as $L(H)+H$, as functions of H_1, denoted as H, for $q_2^H = q_2^L = 0.9$, α=0.5 and different values of $q_1^H = q_1^L = q_1$.



Picture 2

Regions of the primitives where $L = \overline{L_1}(1)$ takes particular values. Here the space of variable primitives is $(\alpha, q_1) = [0,1] \times [0.5,1]$ with $q_1^H = q_1^L = q_1$, and $q_2^H = q_2^L = 0.9$.

Summarizing, imperfect information transmission between selectors limits the number of low signals selected in stage 1. The mixing effect gets relatively stronger if: (i) the prior share of high type alternatives in the population is larger; (ii) signaling in stage 1 is more informative; (iii) the number of high signals in stage 1 is larger.

Having established the properties of the equilibrium in the case of no memory, we turn to another imperfection of information processing, namely to imperfect information acquisition. This is information overload case.

4.3 Limited attention

In the perfect information environment case, we have seen that the preferences of selector 2 are different for different values of the primitives of the model. If the accuracy in stage 1 is higher than in stage 2, signal realization $\left(x_1^H, x_2^L\right)$ is preferred to $\left(x_1^L, x_2^H\right)$ and the other way around. Under limited attention, however, the accuracy in stage 2 is endogenously determined by the sample size, i.e., by the number of alternatives selected in stage 1. On the other hand, the sample size is the variable that is readily observable in stage 2. Hence, selector 1, by selecting alternatives, implicitly selects one of the two possible preference relations, and selector 2 has consistent preferences in all states of the world, i.e., for all possible signal realization in stage 2.

Thus, like in perfect information environment case, selector 2 has a unique weakly undominated strategy, which is determined by its preference relation, which, in turn, is determined by selector 1. It turns out that there exists a generically unique PBE of the game, which is stated in the following proposition.

Proposition 3.3 *In the limited attention case, there exists a generically unique PBE such that:*

a) There exists an upper-bound $\bar{H}_1 \in [1, \infty)$ such that $h_1^* (H_1, L_1) \leq \bar{H}_1$, i.e., player 1 selects not more than $\bar{H}_1$ high signals.

b) For any $\bar{H}_1 \geq 0$ there exists an upper-bound $\bar{L}_1 (H_1) \in [0, \infty)$ and a finite set of lower-bounds $\left\{\underline{L}_1^k\right\}_{k=1}^K$, $1 \leq K < \bar{L}_1 (H_1)$, $0 \leq \underline{L}_1^1 \leq \underline{L}_1^k \leq \underline{L}_1^{k+1} \leq \bar{L}_1 (H_1)$, such that:

$$l_1^* (H_1, L_1) = \begin{cases} \min\left\{L_1, \bar{L}_1 (H_1)\right\} & \text{if } L_1 \geq L_1^K (H_1) \\ \underline{L}_1^k & \text{if } L_1 \in [\underline{L}_1^k, \underline{L}_1^{k+1}) \end{cases}$$

i.e., if the number of low signals does not exceed $\underline{L}_1^{k+1}$, only $\underline{L}_1^k$ of them are selected; otherwise all of them up to $\bar{L}_1 (H_1)$ are selected in stage 1.

c) $h_1^* (H_1, L_1)$ is a weakly increasing function of H_1 and does not depend on L_1.

d) $\bar{L}_1 (H_1) = 0$ whenever $H_1 > \bar{H}_1$.

e) $l_1^*(h, L_1) = \underline{L}_1^1(h) = \bar{L}_1(h) = 0$ *for all* $h \geq h_1^*$ *if* $q_2(h_1^*) < q_1$*, i.e., if selecting optimal number of high signals* h_1^* *makes the screening accuracy in stage 2 lower than in stage 1, none of low signals are selected in stage 1.*

f) Player 2 selects an alternative in accordance with the following preference relation:

$\left(x_1^H, x_2^H\right) \succ \left(x_1^L, x_2^H\right) \succ \left(x_1^H, x_2^L\right) \succ \left(x_1^L, x_2^L\right)$, *if* $q_2(H_2 + L_2) > q_1$

$\left(x_1^H, x_2^H\right) \succ \left(x_1^H, x_2^L\right) \succ \left(x_1^L, x_2^H\right) \succ \left(x_1^L, x_2^L\right)$, *if* $q_2(H_2 + L_2) \leq q_1$

The statements of Proposition 3.3 can be understood as follows. First, like in Proposition 3.2, PBE is generically unique as the pay-off function of player 1 turns out to be a regular analytical function of the model's primitives and, therefore, takes generically different values for all different values of its arguments. Then, selecting one extra signal has two effects on the pay-off. The first effect is the same sample size effect as in Proposition 3.2: selecting more signals in stage 1 increases the probability of observing at least one high signal in stage 2. The other effect, which is now a decreasing accuracy effect, is negative due to the decrease in q_2. The decreasing accuracy effect, in contrast to the mixing effect from Proposition 3.2, prevents selecting too many signals of both types, thus, there are upper-bounds $\bar{H}_1$ and $\bar{L}_1$, parts (a) and (b). If only a few low signals are available, it might be possible that taking none of them is optimal even if $\bar{L}_1 > 0$. Thus, the existence of the lower bound $\underline{L}_1$ is established.

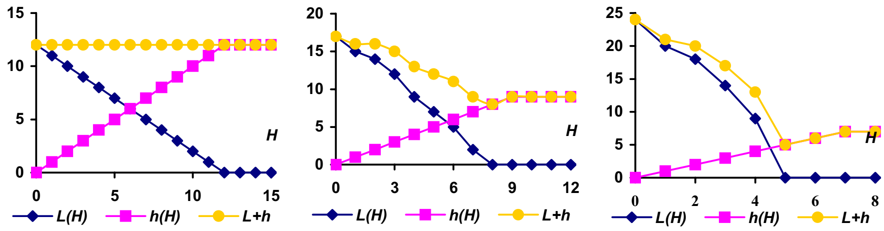
High signals in stage 1 are always more favorable than low signals and, therefore, $h_1^*(H_1, L_1)$ is a weakly increasing function of H_1 and does not depend on L_1, part (c). Due to the same reason, none of low signals will be selected if some high signals are neglected, part (d). Selector 1 selects low signals in addition to high signals only if the screening accuracy in stage 2 is still higher than in stage 1. Otherwise none of low signal will be selected, part (e).

Comparing Proposition 3.2 and Proposition 3.3 one may note that the only difference between no memory case and information overload case is when signals in stage 1 are homogeneous: in information overload case the number of alternatives to be selected is bounded whereas in no memory case it is not. In order to show other differences between the two cases, we provide numerically calculated functions $\bar{L}_1(H_1)$ and $h_1^*(H_1)$ in the following example.

Example 2 *In this example, we have selected a function* $q_2(n) = 0.89 + 0.01 \cdot \frac{2}{n}$*for modeling the decreasing accuracy in order to make it comparable with Example 1. In both cases,* $q_2(2) = 0.9$ *and the decrease in* q_2 *here seems to be negligible. It turns out, however, that even such small decrease in stage 2 accuracy is enough to ensure that only a few alternatives are passed to stage 2. Intuitively, even for relatively small numbers of selected* *alternatives, the decreasing accuracy effect, which vanishes at a polynomial rate, dominates the sample size effect, which vanishes exponentially.*

Picture 3 shows first that, as in no communication case, the upper-bound $\bar{L}_1(H_1)$ *is a non-increasing function and is strictly decreasing function when* $\bar{L}_1(H_1) > 0$. *Second, the maximum sample size in stage 1 is not a monotone function.*

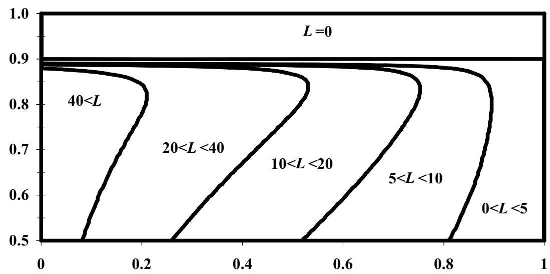
Like in no communication case, the prior α *monotonically affects the sample composition: the higher the prior is, the less low signals are selected, see Picture 4. In contrast, the accuracy in stage 1 affects the sample composition non-monotonically: the minimum number of low signals is selected, either for very informative stage 1 signals, or for almost uninformative ones. Contrary to the no communication case, under limited attention the decrease in stage 2 accuracy prevents selector 1 from taking very many low signals even if stage 1 signaling is almost uninformative.*



a) $q_1 = 0.51$ b) $q_1 = 0.65$ c) $q_1 = 0.78$

Picture 3

The upper-bound $\overline{L_1}(H_1)$, denoted as L, the optimal number of high signals $h_1^*(H_1)$, denoted as h and the maximum sample size in stage 1 L+h as functions of H_1, denoted as H for α=0.5, different values of $q_1^H = q_1^L = q_1$ and $q_2(n) = 0.89 + 0.01 \cdot \frac{2}{n}$.



Picture 4

Regions of the primitives where $L = \overline{L_1}(1)$ takes particular values. Here the space of variable primitives is $(\alpha, q_1) = [0,1] \times [0.5,1]$ with $q_1^H = q_1^L = q_1$, and $q_2(n) = 0.89 + 0.01 \cdot \frac{2}{n}$.

As we have seen, both types of information processing imperfections yield inefficiency. The causes of the inefficiency, however, are different. In no communication case, it is the purely statistical mixing effect that reduces the incentives to mix heterogeneous signals into a single pool. In the limited attention case, it is the decrease in accuracy that makes selection of large samples too costly. Example 2 shows that even a tiny decrease in accuracy significantly affects the optimal behavior of the network. In order to highlight the common features and the distinctions of these two effects, we compare equilibrium properties obtained analytically in Proposition 3.2 and Proposition 3.3, and obtained numerically in Example 1 and Example 2.

First, the mixing effect manifests itself only in heterogeneous samples. That is why in no communication case no information is discarded if only high or only low signals are observed. Furthermore, when the accuracy in stage 1 vanishes, all alternatives will be selected. This is so because the accuracy in stage 1 determines how heterogeneous high and low signals are.

When the signaling in stage 1 is very accurate, but still less accurate than in stage 2, the sample size effect, which is the other determinant of the optimal selection rule, vanishes. In this case, the limited attention prevents selecting low signals in stage 1 whatsoever. Therefore, and this is the second principal difference between lack of communication and the limited attention, the accuracy in stage 1 affects the number of low signals selected non-monotonically in the no communication case whereas in the limited attention case this dependence in monotone.

Apart from these two distinctions, all the other equilibrium features of lack of communication and limited attention are very much alike due to the similarities between the mixing effect and the decreasing accuracy effect. Both effects become stronger relative to the sample size effect when the number of high signals increases. That is why the number of low signals selected weakly decreases with the number of observed high signals and weakly increases with the number of observed low signals in both settings. When the number of high signals is sufficiently large, none of low signals will be selected.

Both effects become stronger also for large values of the prior. In other words, the inefficiency due to informational imperfections is the highest when there are only few low types in the population. On the contrary, when the initial share of high types is very low and, therefore, both high and low signals in stage 1 came from low types almost surely, both effects vanish and so does the inefficiency. The last, but not least, common feature of the lack of communication and the limited attention scenarios is that the resulting sample size in stage 1 is a non-monotone function of the sample composition. This becomes very important in a multi-stage generalization of the model.

4.4 Lack of communication and limited attention

We finish the analysis of the general 2-stage binary selection model by allowing both sources of imperfections. Combining Proposition 3.2 and Proposition 3.3 we get the following result.

Proposition 3.4 *In the no communication and lack of information case, there exists a generically unique PBE such that:*

a) There exists an upper-bound $\bar{H}_1 \in [1, \infty)$ *such that* $h_1^*(H_1, L_1) \le \bar{H}_1$*, i.e., player 1 selects not more than* $\bar{H}_1$ *high signals.*

b) For any $\bar{H}_1 \ge 0$ *there exists an upper-bound* $\bar{L}_1(H_1) \in [0, \infty)$ *and a finite set of lower-bounds* $\left\{\underline{L}_1^k\right\}_{k=1}^K$, $1 \le K < \bar{L}_1(H_1)$, $0 \le \underline{L}_1^1 \le \underline{L}_1^k \le \underline{L}_1^{k+1} \le \bar{L}_1(H_1)$, *such that:*

$$l_1^*(H_1, L_1) = \begin{cases} \min\left\{L_1, \bar{L}_1(H_1)\right\} & \text{if } L_1 \ge L_1^K(H_1) \\ \underline{L}_1^k & \text{if } L_1 \in [\underline{L}_1^k, \underline{L}_1^{k+1}) \end{cases}$$

i.e., if the number of low signals does not exceed $\underline{L}_1^{k+1}$*, only* $\underline{L}_1^k$ *of them are selected; otherwise all of them up to* $\bar{L}_1(H_1)$ *are selected in stage 1.*

c) $\bar{L}_1(H_1)$ *does not increase and strictly decreases whenever* $\bar{L}_1(H_1) > 0$.

d) There exists a number $\tilde{H}_1 \le \bar{L}_1(1) + 1$ *such that* $l_1^*(h, L_1) = \underline{L}_1(h) = \bar{L}_1(h) = 0$ *for all* $h > \tilde{H}_1$*, i.e., if there are sufficiently many high signals, none of low signals are selected in stage 1.*

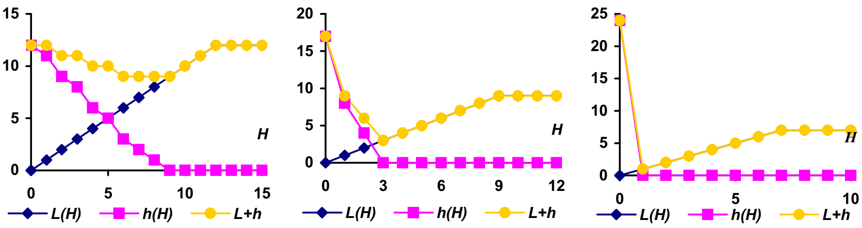
e) $l_1^*(h, L_1) = \underline{L}_1^1(h) = \bar{L}_1(h) = 0$ *for all* $h \ge h_1^*$ *if* $q_2(h_1^*) < q_1$*, i.e., if selecting optimal number of high signals* h_1^* *makes the screening accuracy in stage 2 lower than in stage 1, none of low signals are selected in stage 1.*

f) The total sample size in stage 1 $h_1^* + l_1^*$ *is not a strictly monotone function of* H_1*: it weakly decreases when* $l_1^* > 0$ *and strictly increases when* $l_1^* = 0$.

g) Player 2 selects an alternative in accordance with the following preference relation (1).

The proof of Proposition 3.4 can be obtained by adjusting the proof of Proposition 3.2 to decreasing function $q_2(n)$ and, therefore, is omitted. Naturally, when both types of informational imperfections are present in the model, the inefficiency is the largest. The following Example 3 shows the result of imposing simultaneously no communication and limited attention.

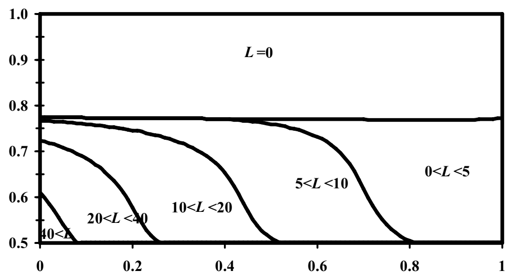
Example 3 *This example differs from Example 2 only in the communication assumption. Picture 5 shows that, in the no communication with limited attention case, the sample size in stage 2 is even smaller than in the limited attention case. Next, Picture 6 shows that the regions of the prior* α *and the first stage accuracy* $q_1^H = q_1^L$ *where selector 1 selects many low signals get smaller. One may see that there hardly can be found any criteria that* *would allow us to classify both types of imperfections based on exogenous variables only. As we have already noticed, the major difference between no communication and limited attention is in the way they react to changes in the screening accuracy of stage 1, compare Picture 2, Picture 4 and Picture 6.*



a) $q_1 = 0.51$ b) $q_1 = 0.65$ c) $q_1 = 0.78$

Picture 5.

The upper-bound $\overline{L_1}(H_1)$, denoted as L, the optimal number of high signals $h_1^*(H_1)$, denoted as h and the maximum sample size in stage 1 $L+h$ as functions of H_1, denoted as H for α=0.5, different values of $q_1^H = q_1^L = q_1$ and $q_2(n) = 0.89 + 0.01 \cdot \frac{2}{n}$.



Picture 6.

Regions of the primitives where $L = \overline{L_1}(1)$ takes particular values. Here the space of variable primitives is $(\alpha, q_1) = [0,1] \times [0.5,1]$ with $q_1^H = q_1^L = q_1$, and $q_2(n) = 0.89 + 0.01 \cdot \frac{2}{n}$.

We have seen that, whatever informational imperfections are, some potentially valuable information is discarded. Under no communication then only low signals will be neglected due to the mixing effect. Under limited attention, both types are affected due to the decreasing accuracy effect. The resulting sample size in stage 2 turns out to be a non-monotone function of the signal composition in stage 1. Under no communication, this makes the model extremely difficult for analytical analysis if there are more than two selection stages. Indeed, Bayesian updates of the beliefs about signaling history of each alternative selected in stage 1 requires accounting for a bi-variate binomial distribution of signal realizations in stage 2, a 4-variate distribution in stage 2 and, in general $2t-1$-variate distribution in stage t.

In the following section, we will show a way to organize stage-screening procedures such that the model becomes analytically tractable in a multistage environment.

5 Filtering selection

In the previous section, I have analyzed the selection game for the case in which both types of alternatives can produce either a high or a low signal. It turns out that imposing stronger requirements on the types of signal outcomes that each alternative can generate, allows me to generalize the model to an arbitrary number of selection stages. In each stage t of selection, player t observes signals of all n_t previously selected alternatives in accordance with revealing probabilities $q_t^H = \Pr\left(x_t^H \mid \theta_H\right)$, $q_t^L = \Pr\left(x_t^L \mid \theta_L\right)$,: H_t high signals and L_t low signals, in total $H_t + L_t = n_t$. He selects a number h_t of high signals and a number l_t of low signals, and passes them to the next stage. In the last stage T, selector T selects one alternative and each player gets the same pay-off, which is equal to the probability that a high type alternative is finally selected.

Assuming $q_t^L = 1$ implies that low type alternatives always generate low signals and, thus, the selection strategy must select all high signals as they have necessarily come from high types. Assuming $q_t^L = 1$ corresponds to a situation in which the screening requirements are set so high that no low quality alternative can ever meet them and therefore I refer to this assumption as high-standard filtering. If, in contrast, $q_t^H = 1$ the selection strategy must filter out all alternatives that have generated low signals as they have necessarily come from low types. This assumption is reminiscent of a situation in which the screening requirements are so low that every high quality alternative meets them for sure and I therefore call this case low-standard filtering selection.

5.1 High standard filtering

It is easily seen that, if at a certain stage t the selector t observes some number of high signals, in a PBE he selects one high signal at random and effectively ends the selection with the pay off of 1. If, on the other hand, he observes only low signals, he can select only low signals but not high signals. Thus, mixing of types does not occur in equilibrium, which has a great impact on the solvability of the game with an arbitrary number of stages. Indeed, without mixing effect, the extent available communication plays no role as every player, having observed more than one alternative, infers that all of them generated only low signals in the past. In this subsection I therefore assume that the information-processing limitation takes the form of limited attention only.

It turns out that no information is discarded in a high standard filtering selection procedure, provided the initial set of alternatives is large enough.

Proposition 3.5 *There exists a threshold level of the population size $\bar{N}$ such that for all $N > \bar{N}$ the filtering selection game with $q_t^L = 1$ has a unique PBE $\{(h_t^*, l_t^*)\}_{t=1}^T$such that for all $t = 1, .., T$.*

a) $h_t^* = sign\,(H_t)$, i.e., player t selects one high signal in stage t if there is one, and the team gets the pay-off of 1;

b) $l_t^* = (1 - sign\,(H_t))\,L_t$ and $l_T^* = (1 - sign\,(H_t))$, i.e., player t selects no low signals if there is at least one high signal, and he selects all $L_T = N$ low signals otherwise.

The proof of Proposition 3.5 is in the appendix. When the sample size asymptotically increases, the probability of observing a high signal in the next stage approaches 1. Thus, if the initial set of alternatives is large enough, every selector selects the whole population if no alternative has passed the filter, i.e., only low signals have been generated.

5.2 Low standard filtering

Clearly, if there are high signals available, it is strictly dominated to take any number of low signals in addition to the high signals. The only possibility for selecting low signals is when no high signals are available. In this case, selecting any numbers of alternatives is pay-off equivalent and generates zero pay-off. Thus, the uniqueness property of PBE fails. In what follows I assume that a player who observes only low signals is forced to select only one alternative, effectively ending the selection. Like in the previous case, mixing of types does not occur and therefore communication plays no role. Thus, I will impose here only limited attention.

Contrary to the case in which $q_t^L = 1$, assuming $q_t^H = 1$ always puts a bound on the number of high signal to be selected in a two-stage filtering selection game.

Proposition 3.6 *In the filtering selection game, with $q_1^H = q_2^H = 1$ there exists a generically unique PBE such that:*

a) There exists an upper-bound $\bar{H}_1 \in [0, \infty)$such that ,$h_1^*\,(H_1, L_1) \leq \bar{H}_1$ i.e., player 1 selects not more than $\bar{H}_1$ high signals.

b) $l_t^* = 1 - sign\,(H_t)$, i.e., none of low signals is selected unless there are no high signals available.

c) $h_1^*\,(H_1, L_1)$is a weakly increasing function of H_1 and does not depend on L_1.

d) Player 2 selects an alternative in accordance with the preference relation (1).

The proof of Proposition 3.6 can be easily obtained from the proof of Proposition 3.3 by taking $q_1^H = q_2^H = 1$ and, therefore, is omitted. More interesting results, however, can be obtained when the initial prior α is close enough to 1. The following proposition states the result.

Proposition 3.7 *There exists a threshold level of the prior $\alpha^* < 1$ such that for all $\alpha \in (\alpha^*, 1)$ a T-stage filtering selection game with $q_t^H = 1$ has a unique PBE $\{(h_t^*, l_t^*)\}_{t=1}^T$ such that for all $t = 1, .., T$.*

a) $h_t^* = \min\{2, H_t\}$, $h_T^* = \min\{1, H_T\}$, i.e., player t selects not more than two high signals in stage t;

b) $l_t^* = 1 - sign(H_t)$, i.e., player t selects no low signals if there are high signals available. Otherwise, he selects one low signal and the team gets the pay-off of zero.

We have seen in section 4 that, in the general two-stage selection model, the number of discarded alternatives increases when the prior gets larger. Proposition 3.7 shows to what extent the decreasing accuracy effect (due to limited memory), whatever small it is, limits the number of high signals selected: the minimum possible number of alternatives for making a nontrivial choice in later stages, namely two, will be selected. This result holds true for any strictly decreasing function , even for those that violate Assumption 3. The reason is that the sample size effect vanishes when α approaches 1 and only decreasing accuracy effect is still working. The following corollary is a direct consequence of Proposition 3.7.

Corollary 3.1 *If the number of selection stages T in the filtering selection game with $q_t^H = 1$ is sufficiently large, then, for any prior α, starting from a certain stage T^* every selector selects at most 2 high signals, i.e., $h_t^* = \min\{2, H_t\}$ for all $t > T^*$.*

Indeed, selecting only high signals in the beginning of the game assures that at stage T^* the prior share of high types α_T becomes sufficiently close to 1 as

$$\alpha_T = \frac{\alpha}{\alpha + (1-\alpha)\prod_{k=1}^{t}\left(1 - q_k^L(n_k)\right)}$$

and $\lim_{t\to\infty} \alpha_t = 1$. When α does not satisfy the conditions of Proposition 3.7, the number of selected high signals remains to be a relatively small integer. For instance, for two-stage selection game, if the revealing probability function $q_t^L(n)$satisfies the following condition for all $n > \bar{n}$:

$$q_2^L(n) - q_2^L(n+1) > (1-\alpha)^{n-1}\left(q_2^L(n+1)\right)^n \tag{2}$$

then player 1 never selects more than $\bar{n}$ high signals[4]. One can easily see that condition (2) is satisfied for $n \geq 2$ for any strictly decreasing function $q_2^L(n)$ when α approaches 1, which is exploited in the proof of Proposition 3.7. Condition (2) can also be generalized for an arbitrary number of selection stages. Due to the exponential structure of the right-hand-side of (2), the condition is satisfied for relatively small numbers. For example, even for the tiny decrease in the accuracy generated by the function $q_2^L(n) = 0.9 + 0.000001\frac{1}{n}$and for $\alpha = 0.5$ the inequality is satisfied for all $n \geq 27$ and, therefore, not more than 27 alternatives are selected. The true upper-bound in this case is equal to 25.

6 Conclusion

I have developed a hierarchical model of rational agents with internal informational imperfections. Those imperfections are introduced by assuming limits on information acquisition (limited attention) and information transmission (lack of communication) that are usually considered separately in the literature. In this framework, I am able to challenge the paradigm «more information is better» as in my model neglecting valuable information emerges as an endogenous behavior of rational agents. The forces and mechanisms responsible for such optimal behavior are also investigated in deep details. In case of limited attention it is the indirect cost of processing large samples that prevents a rational agent from passing too many alternatives to the next selection stages. In case of lack of communication neglecting potentially valuable information is optimal only when the agent observes heterogeneous signal realizations: mixing different signals in a single pool to be passed to the next stage makes it impossible to distinguish between them later on and, therefore, decreases the probability of selecting the best alternative.

Results that are even more striking are obtained, when we turn our attention to screening procedures that take a form of filtering. When selection requirements at all stages are weak, meaning that good alternatives always satisfy them and bad alternatives are gradually filtered out, a very large amount of information must be neglected: relatively few alternatives are sufficient in order to make an efficient choice. The opposite case, when selection requirements at each stage are highly demanding, meaning that bad alternatives never satisfy them and even good alternatives may fail to do so, is the only example where information is never discarded. For any values of the prior and revealing probabilities, either a surely high type alternative is selected, or the whole set of alternatives that did not meet the requirement is passed to the next selection stage.

[4]See appendix for the proof.

7 Appendix to Chapter 3

Proof of Proposition 3.2.

First, we derive the team's pay-off function $u(h_1, l_1)$ provided selector 2 plays his unique weakly undominated strategy and the screening accuracy in stage 1 is lower than in stage 2. $u(\cdot,\cdot)$ turns out to be a rational analytical function of the model primitives, $\alpha, q_1^H, q_1^L, q_2^H, q_2^L$ and, therefore, it takes generically different values for different values of its arguments $(h_1, l_1) \in [0, H_1] \times [0, L_1]$. Hence, there exists a generically unique PBE. Then, we show that $u(\cdot,\cdot)$ strictly increases with h_1 and, therefore, $h_1^*(H_1, L_1) = H_1$, i.e., statement (a) of the proposition. Next, we fix $H_1 \geq 1$ and investigate the shape of $u(\cdot,\cdot)$ as a function of discrete argument L_1. It turns out that $u(\cdot,\cdot)$ may generically have two local maxima: the interior maximum at $l_1 > 0$ and the corner maximum at $l_1 = 0$. We define an upper-bound as the value of l_1 at which $u(\cdot,\cdot)$ attains its global maximum:

$$\bar{L}_1(H_1) = \arg\max_{0 \leq l_1} u(H_1, l_1)$$

If the interior maximum does not exist we define $\bar{L}_1(H_1) = 0$ as in this case $u(\cdot,\cdot)$ strictly decreases for all l_1. Then, we define a lower-bound as the smallest l_1 that yields at least $u(H_1, 0)$:

$$\underline{L}(H_1) = \begin{cases} 0, & \text{if } u(H_1, l_1) < u(H_1, 0) \text{ for all } l_1 > 0 \\ \min\left\{l_1 \mid 0 < l_1 \leq \bar{L}_1(H_1), u(H_1, l_1) \geq u(H_1, 0)\right\}, & \text{otherwise} \end{cases}$$

It is easy to see that the optimal number of low signals l_1^* that selector 1 has to select, which is defined as $l_1^*(H_1, L_1) = \arg\max_{0 \leq l_1 \leq L_1} u(H_1, l_1)$, is zero if $L_1 < \underline{L}_1$ and is equal to $\min\left\{L_1, \bar{L}_1(H_1)\right\}$ if $L_1 \geq \underline{L}_1$, i.e., statement (b) of the proposition. Lastly, we derive the properties (c), (d), (f) and (g) of the functions $\bar{L}_1(H_1)$ and $\underline{L}_1(H_1)$. In order to prove part (e) we note that if the screening accuracy in stage 1 is higher than in stage 2, none of low signals are selected in stage 1 in the presence of high signals, that is $l_1^*(h, L_1) = \bar{L}_1(h) = \underline{L}_1(h) = 0$. Part (h) is trivial. In what follows we use the following notations:

$$\phi_{LL} = \gamma_1^L\left(1 - q_2^H\right) + \left(1 - \gamma_1^L\right) q_2^L$$

$$\phi_{HL} = \gamma_1^H\left(1 - q_2^H\right) + \left(1 - \gamma_1^H\right) q_2^L$$

$$\gamma_{HH} = \frac{\gamma_1^H q_2^H}{1 - \phi_{HL}}$$

$$\gamma_{LH} = \frac{\gamma_1^L q_2^H}{1 - \phi_{LL}}$$

$$\gamma_{HL} = \frac{\gamma_1^H \left(1 - q_2^H\right)}{\phi_{HL}}$$

$$\gamma_{LL} = \frac{\gamma_1^L \left(1 - q_2^H\right)}{\phi_{LL}}$$

Suppose player 1 selects h_1 high signals and l_1 low signals . For any x and y such that $0 \leq x \leq h_1$ and $0 \leq y \leq l_1$, there is a chance that exactly x alternatives out of h_1 and exactly y alternatives out of l_1 will generate high signals . The probability of this event is given by the following bi-variate binomial distribution:

$$\begin{aligned}
\Pr(x, y \mid h_1, l_1) &= \Pr(x \mid h_1) \Pr(y \mid h_1) \\
&= \binom{x}{h_1} (1 - \phi_{HL})^x (\phi_{HL})^{h_1 - x} \binom{y}{l_1} (1 - \phi_{LL})^y (\phi_{LL})^{l_1 - x}
\end{aligned}$$

When this event occurs receiver 2 observes $x + y$ high signals. If $x + y > 0$ the pay-off of the receivers is $\frac{x}{x+y}\gamma_{HH} + \frac{y}{x+y}\gamma_{LH}$. If, on the other hand, $x = y = 0$, the pay-off is $\frac{h_1}{h_1+l_1}\gamma_{HH} + \frac{l_1}{h_1+l_1}\gamma_{LH}$. Thus, the team's pay-off is:

$$\begin{aligned}
u(h_1, l_1) &= \sum_{\substack{x=y=0 \\ x+y>0}}^{\substack{x=h_1 \\ y=l_1}} \Pr(x, y \mid h_1, l_1) \left(\frac{x}{x+y}\gamma_{HH} + \frac{y}{x+y}\gamma_{LH} \right) + \\
&\quad + \Pr(0, 0 \mid h_1, l_1) \left(\frac{h_1}{h_1 + l_1}\gamma_{HL} + \frac{l_1}{h_1 + l_1}\gamma_{LL} \right) \\
&= \gamma_{LH} + (\gamma_{HH} - \gamma_{LH}) \sum_{\substack{x=1 \\ y=0}}^{\substack{x=h_1 \\ y=l_1}} \Pr(x, y \mid h_1, l_1) \frac{x}{x+y} + \\
&\quad + \Pr(0, 0 \mid h_1, l_1) \left(\frac{h_1}{h_1 + l_1} (\gamma_{HL} - \gamma_{LL}) - (\gamma_{HL} - \gamma_{LL}) \right)
\end{aligned}$$

Converting the finite sum above into an integral yields:

$$\begin{aligned}
u(h_1, l_1) &= \gamma_{HL} + (\gamma_{HH} - \gamma_{LH}) \times \\
&\quad \times \int_{w=0}^{w=1} (\phi_{LL} + (1 - \phi_{LL}) w)^{l_1} \, d(\phi_{HL} + (1 - \phi_{HL}) w)^{h_1} \, dw + \\
&\quad + (\phi_{HL})^{h_1} (\phi_{LL})^{l_1} \left(\frac{h_1}{h_1 + l_1} (\gamma_{HL} - \gamma_{LL}) - (\gamma_{LH} - \gamma_{LL}) \right)
\end{aligned}$$

It is a routine to see that $u(h_1+1, l_1)$ can be written as:

$$\begin{aligned}
u(h_1+1, l_1) &= (1-\phi_{HL}) \sum_{\substack{x=y=0}}^{\substack{x=h_1\\y=l_1}} \Pr(x, y \mid h_1, l_1) \times \\
&\times \left(\frac{x+1}{x+1+y} \gamma_{HH} + \frac{y}{x+1+y} \gamma_{LH} \right) + \\
&+\phi_{HL} \sum_{\substack{x=y=0\\x+y>0}}^{\substack{x=h_1\\y=l_1}} \Pr(x, y \mid h_1, l_1) \left(\frac{x}{x+y} \gamma_{HH} + \frac{y}{x+y} \gamma_{LH} \right) + \\
&+\phi_{HL} \Pr(0, 0 \mid h_1, l_1) \left(\frac{h_1+1}{h_1+1+l_1} \gamma_{HL} + \frac{l_1}{h_1+1+l_1} \gamma_{LL} \right)
\end{aligned}$$

and, then, that $\Delta_h u(h_1, l_1) \equiv u(h_1+1, l_1) - u(h_1, l_1) > 0$:

$$\begin{aligned}
\Delta_h u(h_1, l_1) &= \Pr(0, 0 \mid h_1, l_1)(1-\phi_{HL})(\gamma_{HH} - \gamma_{HL}) + \\
&+ \Pr(0, 0 \mid h_1, l_1) \frac{l_1(\gamma_{HL} - \gamma_{LL})}{h_1+1+l_1} \left((1-\phi_{HL}) + \frac{1}{h_1+l_1} \right) + \\
&+ (1-\phi_{HL})(\gamma_{HH} - \gamma_{LH}) \times \\
&\times \sum_{\substack{x=y>0}}^{\substack{x=h_1\\y=l_1}} \Pr(x, y \mid h_1, l_1) \frac{y}{(x+1+y)(x+y)} \\
&> 0
\end{aligned}$$

Thus, $h_1^*(H_1, L_1) = H_1$. Let us now fix any and $h_1 \geq 1$ and define

$$\begin{aligned}
D(h_1, l_1) &\equiv \frac{\Delta_l u(h_1, l_1)}{h_1 (\phi_{HL})^{h_1} (\phi_{LL})^{l_1} (\gamma_{HL} - \gamma_{LL})} \\
&\equiv \frac{u(h_1, l_1+1) - u(h_1, l_1)}{h_1 (\phi_{HL})^{h_1} (\phi_{LL})^{l_1} (\gamma_{HL} - \gamma_{LL})} \\
D(h_1, l_1) &= \left(\frac{\gamma_{HH} - \gamma_{LH}}{\gamma_{HL} - \gamma_{LL}} \right) \frac{(1-\phi_{HL})(1-\phi_{LL})}{\phi_{HL}} \times \\
&\times \int_{w=0}^{w=1} (1-w) \left(1 + \frac{1-\phi_{LL}}{\phi_{LL}} w \right)^{l_1} \left(1 + \frac{1-\phi_{HL}}{\phi_{HL}} w \right)^{h_1-1} dw + \\
&+ \frac{\phi_{LL}}{h_1+1+l_1} - \frac{1}{h_1+l_1} + \left(\frac{\gamma_{LH} - \gamma_{LL}}{\gamma_{HL} - \gamma_{LL}} \right) \left(\frac{1-\phi_{LL}}{h_1} \right)
\end{aligned}$$

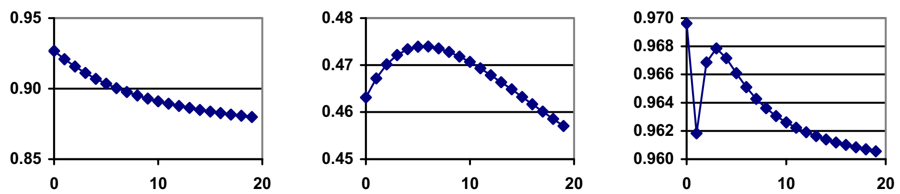
and

$$
\begin{aligned}
E\left(h_1,l_1\right) &\equiv \Delta_l D\left(h_1,l_1\right) \equiv D\left(h_1,l_1+1\right) - D\left(h_1,l_1\right) \\
E\left(h_1,l_1\right) &= \left(\frac{\gamma_{HH}-\gamma_{LH}}{\gamma_{HL}-\gamma_{LL}}\right)\frac{\left(1-\phi_{HL}\right)\left(1-\phi_{LL}\right)^2}{\phi_{HL}\phi_{LL}}\times \\
&\times\int_{w=0}^{w=1}\left(1-w\right)w\left(1+\frac{1-\phi_{LL}}{\phi_{LL}}w\right)^{l_1}\left(1+\frac{1-\phi_{HL}}{\phi_{HL}}w\right)^{h_1-1}dw+ \\
&+\frac{1}{\left(h_1+1+l_1\right)\left(h_1+2+l_1\right)}\left(\left(1-\phi_{LL}\right)+\frac{2}{h_1+l_1}\right)
\end{aligned}
$$

It is easily seen that $\frac{\partial E(h_1,l_1)}{\partial l_1}<0$ and $\lim_{l_1\to\infty}E\left(h_1,l_1\right)=-\infty$. Thus there are two cases.

a) $E\left(h_1,0\right)<0$. In this case $E\left(h_1,l_1\right)<E\left(h_1,0\right)<0$ and, therefore, $D\left(h_1,l_1+1\right)<D\left(h_1,l_1\right)$ and $\lim_{l_1\to\infty}D\left(h_1,l_1\right)=-\infty$. Thus, $D\left(h_1,0\right)$ determines the behavior of $u\left(h_1,l_1\right)$. If $D\left(h_1,0\right)<0$ then $D\left(h_1,l_1\right)<0$ and u always decreases. This happens, e.g., for $\alpha=0.5$, $q_1^H=q_1^L=0.6$, $q_2^H=q_2^L=0.9$, $h_1=6$, See picture 7(a). If ,on the other hand, $D\left(h_1,0\right)>0$ then, first u increases to its interior maximum $\bar{L}_1\left(H_1\right)$ and decreases afterwards. This happens, e.g., for $\alpha=0.1$, $q_1^H=q_1^L=0.6$, $q_2^H=q_2^L=0.9$, $h_1=6$, see picture 7(b).

b) $E\left(h_1,0\right)>0$. In this case there exists a number X such that $E\left(h_1,X\right)<0<E\left(h_1,X-1\right)$, i.e. $D\left(h_1,l_1\right)$ has a unique maximum at $l_1=X$. If $D\left(h_1,X\right)<0$ then $D\left(h_1,l_1\right)<0$ and u always decreases, as in picture 7(a). If, on the other hand, $D\left(h_1,X\right)>0$ then u has a unique interior maximum at $l_1>X$ and decreases afterwards. In this case, if $D\left(h_1,0\right)>0$, u increases for all $l_1<X$, as in Picture 7(b); if $D\left(h_1,0\right)<0$, u has a local minimum for some $l_1<X$. This happens, e.g., for $\alpha=0.9$, $q_1^H=q_1^L=0.78$, $q_2^H=q_2^L=0.9$, $h_1=1$, see picture 7(c).



a) b) c)

Picture 7.

For all three types of shapes of u the definitions of the upper-bound and the lower-bound are consistent, thus, statement (b) of the proposition is proven.

In order to show that is a decreasing function we consider 3 cases.

a) Suppose that for both $h_1 = H_1$ and $h_1 = H_1 + 1$, $u(h_1, l_1)$ attains its global maximum at the interior points $l_1 = \bar{L}_1(H_1) > 0$ and $l_1 = \bar{L}_1(H_1 + 1) > 0$ respectively. We will show that $\bar{L}_1(H_1 + 1) \leq \bar{L}_1(H_1) - 1$. To this end we note that $D(H_1, l_1) < 0$ for all $l_1 \geq \bar{L}_1(H_1)$. Let us consider a difference $F \equiv D(H_1 + 1, l_1 - 1) - D(H_1, l_1)$:

$$\begin{aligned}
F &= \left(\frac{\gamma_{LH} - \gamma_{LL}}{\gamma_{HL} - \gamma_{LL}}\right) \frac{1 - \phi_{LL}}{H_1(H_1 + 1)} - \\
&\quad - \left(\frac{\gamma_{HH} - \gamma_{LH}}{\gamma_{HL} - \gamma_{LL}}\right) \frac{(1 - \phi_{HL})(1 - \phi_{LL})(\phi_{LL} - \phi_{HL})}{(\phi_{HL})^2 \phi_{LL}} \times \\
&\quad \times \int_{w=0}^{w=1} (1 - w) w \left(1 + \frac{1 - \phi_{LL}}{\phi_{LL}} w\right)^{l_1 - 1} \left(1 + \frac{1 - \phi_{HL}}{\phi_{HL}} w\right)^{H_1 - 1} dw \\
&< 0
\end{aligned}$$

as $\phi_{LL} - \phi_{HL} = \left(\gamma_1^H - \gamma_1^L\right)\left(q_2^H + q_2^L - 1\right) > 0$. Thus, $D(H_1 + 1, l_1) < 0$ for all $l_1 \geq \bar{L}_1(H_1) - 1$ and, therefore, $\bar{L}_1(H_1 + 1) \leq \bar{L}_1(H_1) - 1$, i.e., $\bar{L}_1(H_1)$ strictly decreases.

b) Suppose that for $h_1 = H_1 + 1$, $u(h_1, l_1)$ attains its global maximum at the corner $l_1 = \bar{L}_1(H_1 + 1) = 0$. Then, trivially $0 = \bar{L}_1(H_1 + 1) \leq \bar{L}_1(H_1) \geq 0$.

c) The only possibility left is to assume that, for $h_1 = H_1$, $u(h_1, l_1)$ attains its global maximum at the corner $l_1 = \bar{L}_1(H_1) = 0$ while for $h_1 = H_1 + 1 = 0$ it attains its global maximum at the interior point $l_1 = \bar{L}_1(H_1 + 1) > 0$. We will show that this can never be true. As $l_1 = \bar{L}_1(H_1 + 1)$ is assumed to be an interior maximum, it must be that $D(H_1 + 1, l_1) < 0 < D(H_1 + 1, l_1 - 1)$, i.e., $E(H_1 + 1, l_1 - 1) < 0$ for $l_1 = \bar{L}_1(H_1 + 1)$. In addition as, at $h_1 = H_1$, $l_1 = \bar{L}_1(H_1) = 0$ is assumed to be a global maximum, it must be that:

$$\begin{aligned}
G(H_1, l_1) &\equiv \frac{u(H_1, 0) - u(H_1, l_1)}{(\phi_{HL})^{H_1 - 1}(\gamma_{HL} - \gamma_{LL})} \\
&= \frac{\gamma_{HH} - \gamma_{LH}}{\gamma_{HL} - \gamma_{LL}} (1 - \phi_{HL}) H_1 \times \\
&\quad \times \int_{w=0}^{w=1} \left(1 - (\phi_{LL} + (1 - \phi_{LL}) w)^{l_1}\right) \left(1 + \frac{1 - \phi_{HL}}{\phi_{HL}} w\right)^{H_1 - 1} dw + \\
&\quad + \phi_{HL} \left(\left(\frac{\gamma_{HL} - \gamma_{LH}}{\gamma_{HL} - \gamma_{LL}}\right)\left(1 - (\phi_{LL})^{l_1}\right) + \frac{l_1}{H_1 + l_1} (\phi_{LL})^{l_1}\right) \\
&> 0
\end{aligned}$$

for all $l_1 > 0$. But then

$$
\begin{aligned}
Q &\equiv \frac{(1-\phi_{LL})}{l_1\phi_{HL}(\phi_{LL})^{l_1}}\Delta_h G(H_1,l_1) = \frac{1-\phi_{LL}}{l_1\phi_{HL}(\phi_{LL})^{l_1}\left(G(H_1+1,l_1)-F(H_1,l_1)\right)} \\
&= -\frac{1-\phi_{LL}}{(H_1+l_1)(H_1+1+l_1)} + \left(\frac{\gamma_{HH}-\gamma_{LH}}{\gamma_{HL}-\gamma_{LL}}\right)\frac{(1-\phi_{HL})(1-\phi_{LL})}{l_1\phi_{HL}(\phi_{LL})^{l_1}} \times \\
&\quad \times \int_{w=0}^{w=1}\left(\begin{array}{c} \left(1-(\phi_{LL}+(1-\phi_{LL})w)^{l_1}\right)\times \\ \times\left(1+(H_1+1)\frac{1-\phi_{HL}}{\phi_{HL}}w\right)\left(1+\frac{1-\phi_{HL}}{\phi_{HL}}w\right)^{H_1-1} \end{array}\right)dw \\
&> -\frac{1-\phi_{LL}}{(H_1+l_1)(H_1+1+l_1)} + \left(\frac{\gamma_{HH}-\gamma_{LH}}{\gamma_{HL}-\gamma_{LL}}\right)\frac{(1-\phi_{HL})(1-\phi_{LL})^2}{\phi_{HL}\phi_{LL}} \times \\
&\quad \times \int_{w=0}^{w=1}\left(\begin{array}{c} (1-w)\left(1+\frac{1-\phi_{LL}}{\phi_{LL}}w\right)^{l_1-1}\times \\ \times\left(1+\frac{H_1(1-\phi_{HL})w}{\phi_{HL}+(1-\phi_{HL})w}\right)\left(1+\frac{1-\phi_{HL}}{\phi_{HL}}w\right)^{H_1} \end{array}\right)dw \\
&> \frac{2\phi_{LL}}{(H_1+l_1)(H_1+1+l_1)(H_1+2+l_1)} - E(H_1+1,l_1-1) + \\
&\quad + \left(\frac{\gamma_{HH}-\gamma_{LH}}{\gamma_{HL}-\gamma_{LL}}\right)\frac{(1-\phi_{HL})(1-\phi_{LL})^2}{\phi_{HL}\phi_{LL}}\times \\
&\quad \times \int_{w=0}^{w=1}\left(\begin{array}{c} (1-w)\left(1+\frac{1-\phi_{LL}}{\phi_{LL}}w\right)^{l_1-1}\times \\ \left((1-w)+\frac{H_1(1-\phi_{HL})w}{\phi_{HL}+(1-\phi_{HL})w}\right)\left(1+\frac{1-\phi_{HL}}{\phi_{HL}}w\right)^{H_1} \end{array}\right)dw \\
&> 0
\end{aligned}
$$

as $E(H_1+1,l_1-1) < 0$ at $l_1 = \bar{L}_1(H_1+1)$. Therefore, $G(H_1+1,l_1) > 0$ for all $l_1 > 0$ as well. But this contradicts the assumption we made that for, $h_1 = H_1+1$, $u(h_1,l_1)$ attains its global maximum at the interior point $l_1 = \bar{L}_1(H_1+1) > 0$. All the three cases prove part (c) of the proposition. Part (d) follows from

$$u(0,l_1+1) - u(0,l_1) = (\gamma_{LH}-\gamma_{LL})(\phi_{LL})^{l_1}(1-\phi_{LL}) > 0$$

Parts (d) and (g) are direct consequences of part (c). ■

Proof of Proposition 3.3.

First, we derive the team's pay-off function $u(h_1,l_1)$ provided selector 2 plays his unique weakly undominated strategy induced by $q_2(h_1^*) > q_1$. Then, we show that u strictly increases with respect to both arguments provided the accuracy in stage 2 is constant. Thit proves Proposition 3.1 for the case $q_2 > q_1$. When $q_2 < q_1$, the utility is given by the same function $u(h_1,0)$, as no low signals will be selected. This completely proves Proposition 3.1.

As in the proof of Proposition 3.2, $u(h_1, l_1)$ turns out to be a rational analytical function of the model primitives and, therefore, it takes generically different values for different values its arguments . Hence, there exists a generically unique PBE. Next, we show that for any strictly decreasing function $q_2(n)$, which satisfies Assumption 3, $u(h_1, l_1)$ asymptotically decreases with respect to both arguments. This proves the existence of upper bounds $\bar{L}_1$ and $\bar{H}_1$, parts (a) and (b). Then we define lower-bounds $\underline{L}_1^k$ as a convenient way of expressing $l_1^*(H_1, L_1)$ that ends the proof of part (b). Parts (c) and (d) are proven by deriving properties of the upper-bounds. Finally, if $q_2(h_1^*) < q_1$, none of low signals are selected and the team's pay-off function $u(h_1, l_1)$ becomes $u(h_1, 0)$, part (e). In what follows, we use the following notations:

$$\phi_{LL}(n) = \gamma_1^L (1 - q_2(n)) + \left(1 - \gamma_1^L\right) q_2(n)$$

$$\phi_{HL}(n) = \gamma_1^H (1 - q_2(n)) + \left(1 - \gamma_1^H\right) q_2(n)$$

$$\gamma_{HH}(n) = \frac{\gamma_1^H q_2(n)}{1 - \phi_{HL}(n)}$$

$$\gamma_{LH}(n) = \frac{\gamma_1^L q_2(n)}{1 - \phi_{LL}(n)}$$

$$\gamma_{HL}(n) = \frac{\gamma_1^H (1 - q_2(n))}{\phi_{HL}(n)}$$

$$\gamma_{LL}(n) = \frac{\gamma_1^L (1 - q_2(n))}{\phi_{LL}(n)}$$

Having been written without the argument, the above variables are assumed to be evaluated at $q_2 = \underline{q}_2$, or, alternatively at $n \to \infty$. Then, the team's pay-off in case $q_2 > q_1$ is:

$$\begin{aligned} u(h_1, l_1) &= \left(1 - (\phi_{HL})^{h_1}\right) \gamma_{HH} + (\phi_{HL})^{h_1} \left(1 - (\phi_{LL})^{l_1}\right) \gamma_{LH} + \\ &\quad + (\phi_{HL})^{h_1} (\phi_{LL})^{l_1} (\gamma_{LL} + sign(h_1)(\gamma_{HL} - \gamma_{LL})) \\ &= \gamma_{HH} - (\phi_{HL})^{h_1} (\gamma_{HL} - \gamma_{LH}) + (\phi_{LL})^{l_1} \times \\ &\quad \times ((\gamma_{LH} - \gamma_{LL}) - sign(h_1)(\gamma_{HL} - \gamma_{LL})) \end{aligned}$$

It is clearly seen that, for constant q_2, $\partial u / \partial h_1 > 0$ and $\partial u / \partial l_1 > 0$. Thus, in this case, $h_1^* = H_1$ and $l_1^* = L_1$ which proves Proposition 3.1.

When $q_2(n)$ is a strictly decreasing function this is not the case any more. In order to show that we use Assumption 3 that yields:

$$\lim_{n\to\infty} \frac{(\phi_{HL}(n))^n}{q_2(n) - q_2(n+1)} = \lim_{n\to\infty} \frac{(\phi_{LL}(n))^n}{q_2(n) - q_2(n+1)} = 0 \qquad (A.1)$$

First, we show that when only high signals are available at stage 1, i.e., $L_1 = 0$, u asymptotically decreases with h_1. Let $MU(n,0) = u(n,0) - u(n+1,0)$ for $n \geq 1$ denote the marginal disutility of having an extra high signal. Then, using (A.1) yields

$$\lim_{n\to\infty} \frac{MU(n,0)}{q_2(n) - q_2(n+1)} = \lim_{n\to\infty} \frac{\gamma_{HH}(n) - \gamma_{HH}(n+1)}{q_2(n) - q_2(n+1)} = \frac{\gamma_1^H\left(1-\gamma_1^H\right)}{1-\phi_{HL}} > 0$$

Thus, for$L_1 = 0$ there exists an upper bound $\bar{H}_1 < \infty$ such that $h_1^*(H_1, 0) \leq \bar{H}_1$. Then, it is easy to see that $\Delta \equiv u(h+1, n-h-1) - u(h, n-h) > 0$:

$$\begin{aligned}
\Delta &= \gamma_{HH}(n) - (\phi_{HL}(n))^{h+1} \times \\
&\quad \times \Big((\gamma_{HH}(n) - \gamma_{LH}(n)) + (\phi_{LL}(n))^{n-h-1}(\gamma_{LH}(n) - \gamma_{HL}(n))\Big) - \\
&\quad -\gamma_{HH}(n) + (\phi_{HL}(n))^{h} \times \\
&\quad \times \Big((\gamma_{HH}(n) - \gamma_{LH}(n)) + (\phi_{LL}(n))^{n-h}(\gamma_{LH}(n) - \gamma_{HL}(n))\Big) \\
&= (\phi_{HL}(n))^{h}(1-\phi_{HL}(n))(\gamma_{HH}(n) - \gamma_{LH}(n)) + \\
&\quad + (\phi_{HL}(n))^{h}(\phi_{LL}(n))^{n-h-1}\left(\gamma_1^H - \gamma_1^L\right) \times \\
&\quad \times (2q_2(n) - 1)(\gamma_{LH}(n) - \gamma_{HL}(n)) \\
&> 0
\end{aligned}$$

Thus, if $h_1^* < H_1$, i.e., some high signals are neglected, then $l_1^* = 0$, i.e., no low signals will be selected, part (d) of the proposition.

In case $L_1 > 0$ and $h_1^* = H_1 \geq 1$, let $MU(H_1, n-H_1) = u(H_1, n-H_1) - u(H_1, n-H_1+1)$ denote the marginal disutility of having an extra low signal. Then we define:

$$\begin{aligned}
F(H_1) &\equiv \lim_{n\to\infty} \frac{MU(H_1, n-H_1)}{q_2(n) - q_2(n+1)} \\
F(H_1) &\equiv \lim_{n\to\infty} \frac{\gamma_1^H\left(1-\gamma_1^H\right)}{(1-\phi_{HL})^2}\left(1-(\phi_{HL})^{H_1}\right) + \frac{\gamma_1^L\left(1-\gamma_1^L\right)}{(1-\phi_{LL})^2}(\phi_{HL})^{H_1} + \\
&\quad + H_1\frac{\left(2\gamma_1^H - 1\right)\left(\gamma_1^H - \gamma_1^L\right)}{(1-\phi_{HL})(1-\phi_{LL})}\left(1-\underline{q}_2\right)\underline{q}_2(\phi_{HL})^{H_1-1}
\end{aligned}$$

We will show that $F(H_1) > 0$. It is clear that $F(H_1) > 0$ for any $\gamma_1^H \geq 1/2$. In order to show that $F(H_1) > 0$ also for, $\gamma_1^H < 1/2$ we define

$$\begin{aligned}
G(h) &\equiv \frac{(F(h+1) - F(h))(1-\phi_{LL})^2(1-\phi_{HL})}{\left(\gamma_1^H - \gamma_1^L\right)(\phi_{HL})^{h-1}} \\
&= \left(\left(1-\left(\gamma_1^H + \gamma_1^L\right)\right)\left(1-\underline{q}_2\right)^2 - \gamma_1^H\gamma_1^L\left(2\underline{q}_2 - 1\right)\right)\phi_{HL} - \\
&\quad - (\phi_{HL} - h(1-\phi_{HL}))(1-\phi_{LL})\left(1-2\gamma_1^H\right)\left(1-\underline{q}_2\right)\underline{q}_2
\end{aligned}$$

and consider 2 cases:

a) Let be

$$\left(1-\left(\gamma_1^H+\gamma_1^L\right)\right)\left(1-\underline{q}_2\right)^2 \geq \gamma_1^H\gamma_1^L\left(2\underline{q}_2-1\right)+\left(1-\phi_{LL}\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q}_2\right)\underline{q}_2$$

First, G strictly increases as

$$G'(h) = \left(1-\phi_{HL}\right)\left(1-\phi_{LL}\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q}_2\right)\underline{q}_2 > 0$$

Thus, $G(h) \geq G(0)$ where

$$\begin{aligned} G(0) &= \phi_{HL}\left(1-\left(\gamma_1^H+\gamma_1^L\right)\right)\left(1-\underline{q}_2\right)^2 - \phi_{HL}\gamma_1^H\gamma_1^L\left(2\underline{q}_2-1\right) - \\ &\quad -\phi_{HL}\left(1-\phi_{LL}\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q}_2\right)\underline{q}_2 \\ &\geq 0 \end{aligned}$$

In this case $G(h) > 0$ for all $h > 0$. Thus, $F(h+1) - F(h) > 0$. But then for all $H_1 > 0$:

$$F(H_1) > F(0) = \frac{\gamma_1^L\left(1-\gamma_1^L\right)}{\left(1-\phi_{LL}\right)^2}\left(\phi_{HL}\right)^h > 0$$

b) Let be

$$\left(1-\left(\gamma_1^H+\gamma_1^L\right)\right)\left(1-\underline{q}_2\right)^2 < \gamma_1^H\gamma_1^L\left(2\underline{q}_2-1\right)+\left(1-\phi_{LL}\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q}_2\right)\underline{q}_2$$

In this case:

$$\begin{aligned} F(h) &= \frac{\gamma_1^H\left(1-\gamma_1^H\right)}{\left(1-\phi_{HL}\right)^2} - \left(\phi_{HL}\right)^h\left(\frac{\gamma_1^H\left(1-\gamma_1^H\right)}{\left(1-\phi_{HL}\right)^2} - \frac{\gamma_1^L\left(1-\gamma_1^L\right)}{\left(1-\phi_{LL}\right)^2}\right) - \\ &\quad - \left(\phi_{HL}\right)^h h\frac{\left(1-2\gamma_1^H\right)\left(\gamma_1^H-\gamma_1^L\right)}{\left(1-\phi_{HL}\right)\left(1-\phi_{LL}\right)\phi_{HL}}\left(1-\underline{q}_2\right)\underline{q}_2 \\ F(h) &= \frac{\gamma_1^H\left(1-\gamma_1^H\right)}{\left(1-\phi_{HL}\right)^2} - \frac{\left(\gamma_1^H-\gamma_1^L\right)\left(\phi_{HL}\right)^{h-1}}{\left(1-\phi_{HL}\right)^2\left(1-\phi_{LL}\right)^2} \times \\ &\quad \times \left(\begin{array}{c} \left(\left(1-\left(\gamma_1^H+\gamma_1^L\right)\right)\left(1-\underline{q}_2\right)^2 - \gamma_1^H\gamma_1^L\left(2\underline{q}_2-1\right)\right)\phi_{HL} + \\ +h\left(1-2\gamma_1^H\right)\left(1-\phi_{HL}\right)\left(1-\phi_{LL}\right)\left(1-\underline{q}_2\right)\underline{q}_2 \end{array}\right) \\ &> \frac{\gamma_1^H\left(1-\gamma_1^H\right)}{\left(1-\phi_{HL}\right)^2} - \frac{\left(\gamma_1^H-\gamma_1^L\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q}_2\right)\underline{q}_2}{\left(1-\phi_{HL}\right)^2\left(1-\phi_{LL}\right)\phi_{HL}} \times \\ &\quad \times\left(\phi_{HL}+h\left(1-\phi_{HL}\right)\right)\left(\phi_{HL}\right)^h \\ F(h) &> \frac{\gamma_1^H\left(1-\gamma_1^H\right)\left(1-\phi_{LL}\right)\phi_{HL}}{\left(1-\phi_{HL}\right)^2\left(1-\phi_{LL}\right)\phi_{HL}} - \\ &\quad -\frac{\left(\gamma_1^H-\gamma_1^L\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q}_2\right)\underline{q}_2\left(\phi_{HL}+h\left(1-\phi_{HL}\right)\right)\left(\phi_{HL}\right)^h}{\left(1-\phi_{HL}\right)^2\left(1-\phi_{LL}\right)\phi_{HL}} \end{aligned}$$

Let be

$$\begin{aligned} Q(h) &= \gamma_1^H\left(1-\gamma_1^H\right)(1-\phi_{LL})\phi_{HL} - \left(\gamma_1^H-\gamma_1^L\right)\left(1-2\gamma_1^H\right)\times \\ &\quad \times\left(1-\underline{q}_2\right)\underline{q}_2\left(\phi_{HL}+h\left(1-\phi_{HL}\right)\right)\left(\phi_{HL}\right)^h \end{aligned}$$

such that

$$F(h) > \frac{Q(h)}{(1-\phi_{HL})^2(1-\phi_{LL})\phi_{HL}}$$

Then

$$\begin{aligned} Q'(h) &= -\left(\gamma_1^H-\gamma_1^L\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q}_2\right)\underline{q}_2\times \\ &\quad \times\left((1-\phi_{HL})+\left(\phi_{HL}+h(1-\phi_{HL})\right)\ln(\phi_{HL})\right)(\phi_{HL})^h \end{aligned}$$

Using $\ln(x) < x-1$ for $x<1$ yields:

$$\begin{aligned} Q'(h) &> -\left(\gamma_1^H-\gamma_1^L\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q}_2\right)\underline{q}_2\times \\ &\quad \times\left((1-\phi_{HL})+\left(\phi_{HL}+h(1-\phi_{HL})\right)(\phi_{HL}-1)\right)(\phi_{HL})^h \\ &> \left(\gamma_1^H-\gamma_1^L\right)\left(1-2\gamma_1^H\right)(1-\phi_{HL})^2\left(1-\underline{q}_2\right)\underline{q}_2(h-1)(\phi_{HL})^h \geq 0 \end{aligned}$$

Thus,

$$\begin{aligned} Q(h) &> Q(1) = \phi_{HL}\begin{pmatrix} \gamma_1^H\left(1-\gamma_1^H\right)(1-\phi_{LL})- \\ -\left(\gamma_1^H-\gamma_1^L\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q}_2\right)\underline{q}_2 \end{pmatrix} \\ &> \phi_{HL}\left(\gamma_1^H-\gamma_1^L\right)\begin{pmatrix} \left(1-2\gamma_1^H\right)\left(\left(1-\underline{q}_2\right)^2+\gamma_1^L\left(2\underline{q}_2-1\right)\right)+ \\ +\gamma_1^H(1-\phi_{LL}) \end{pmatrix}+ \\ &\quad +\phi_{HL}\gamma_1^L\left(1-\gamma_1^H\right)(1-\phi_{LL}) \\ &> 0 \end{aligned}$$

and, therefore, $F(h) > 0$.

Summarizing both cases yields that $F(H_1) > 0$ for all $H_1 > 0$ and, therefore, $MU(h, n-H_1) > 0$. Thus, for any $h_1^* = H_1 \geq 1$ there exists an upper bound $\bar{L}_1(H_1) < \infty$ such that $l_1^* \leq \bar{L}_1(H_1)$.

If there are no high signals available, i.e., $h_1^* = H_1 = 0$, the marginal disutility of having an extra low signal $MU(0,n) = u(0,n) - u(0,n+1)$, for large n becomes:

$$\begin{aligned} F(0) &\equiv \lim_{n\to\infty}\frac{MU(0,n)}{\underline{q}_2(n)-\underline{q}_2(n+1)} \\ &= \lim_{n\to\infty}\frac{\gamma_{LH}(n)-\gamma_{LH}(n+1)}{q_2(n)-q_2(n+1)} = \frac{\gamma_1^L\left(1-\gamma_1^L\right)}{(1-\phi_{LL})^2} > 0 \end{aligned}$$

That proves the existence of an upper bound $\bar{L}_1(0) < \infty$. The set of lower-bounds is recursively defined as follows:

$$\underline{L}_1^1 = \begin{cases} 0, & \text{if } u(H_1, l_1) < u(H_1, 0) \text{ for all } l_1 > 0 \\ \min\{l_1 \mid 0 < l_1 \leq \underline{L}_1(H_1), u(H_1, l_1) \geq u(H_1, 0)\}, & \text{otherwise} \end{cases}$$

$$\underline{L}_1^{k+1} = \begin{cases} \underline{L}_1^k & \text{if } u(H_1, l_1) < u\left(H_1, \underline{L}_1^k\right) \text{ for all } l_1 > \underline{L}_1^k \\ \min\left\{l_1 \mid \underline{L}_1^k < l_1 \leq \underline{L}_1(H_1), u(H_1, l_1) \geq u\left(H_1, \underline{L}_1^k\right)\right\}, & \text{otherwise} \end{cases}$$

We stop this process at stage K when $\underline{L}_1^{K+1} = \underline{L}_1^K > \underline{L}_1^{K-1}$. It is easy to see that the optimal number of low signals l_1^* that selector 1 has to select, which is defined as $l_1^*(H_1, L_1) = \arg\max_{0 \leq l_1 \leq L_1} u(H_1, l_1)$, is equal to $\underline{L}_1^k$ when $\underline{L}_1^k \leq \underline{L}_1 < \underline{L}_1^{k+1}$, and is equal to $\min\left\{L_1, \bar{L}_1(H_1)\right\}$ when $L_1 \geq \underline{L}_1^K$, that ends the proof of the proposition. ■

Proof of Proposition 3.5.

In what follows, we will use the following notations:

$$\begin{aligned} \phi_t^L(\alpha_t, n_t) &= 1 - \alpha_t q_t^L(n) \\ \gamma_t^L(\alpha_t, n_t) &= \frac{\alpha_t\left(1 - q_t^L(n_t)\right)}{\phi_t^L(\alpha_t, n_t)} \end{aligned}$$

Here α_t stands for the prior at the beginning of stage t; γ_t^L stands for the posterior at the end of stage t provided a low signal is observed, such that $\alpha_t = \gamma_{t-1}^L$; n_t stands for the sample size in the beginning of stage t, such that $n_{t+1} = H_t + 1 + L_{t+1} = h_t + l_t$. If player t observes some number of high signals, the strategy $h_t^* = 1$ weakly dominates all the others. Thus, $h_t^* = sign(H_t)$. The rest of the proof is based on induction assuming that only low signals were available. The ex-ante pay-off function in the last stage T is given by

$$\begin{aligned} \bar{u}_T(\alpha_T, n_T) &= \left(1 - \left(\phi_T^L(\alpha_T, n_T)\right)^{n_T}\right) + \left(\phi_T^L(\alpha_T, n_T)\right)^{n_T} \gamma_T^L \\ &= 1 - (1 - \alpha_T)\left(1 - \alpha_T q_T^L(n_T)\right)^{n_T - 1} \\ &= 1 - (1 - \alpha_T)\left(1 - \alpha_T\left(1 - \prod_{k=T}^{T}\left(1 - q_T^L(n_T)\right)\right)\right)^{n_T - 1} \end{aligned}$$

Suppose that in stage t the ex-ante pay-off function $\bar{u}_t(\alpha_t, n_t)$ is

$$\begin{aligned} \bar{u}_t(\alpha_t, n_t) &= 1 - (1 - \alpha_t)\left(1 - \alpha_t q_t^L(n_t)\right)^{n_t - 1} \\ &= 1 - (1 - \alpha_t)\left(1 - \alpha_t\left(1 - \prod_{k=t}^{T}\left(1 - q_k^L(n_t)\right)\right)\right)^{n_t - 1} \end{aligned}$$

The corresponding reduced form pay-off function $u_{t-1}(\alpha_{t-1}, l_{t-1})$ in stage $t-1$ is given by:

$$\begin{aligned}
u_{t-1}(\alpha_{t-1}, l_{t-1}) &= 1 - \left(1 - \gamma_{t-1}^{L}\right) \times \\
&\times \left(1 - \gamma_{t-1}^{L}\left(1 - \prod_{k=t}^{T}\left(1 - q_k^L(l_{t-1})\right)\right)\right)^{l_{t-1}-1} \\
&= 1 - \frac{1 - \alpha_{t-1}}{1 - \alpha_{t-1} q_{t-1}^{L}(n_{t-1})} \times \\
&\times \left(\begin{array}{c} 1 - \frac{\alpha_{t-1}\left(1 - q_{t-1}^{L}(n_{t-1})\right)}{1 - \alpha_{t-1} q_{t-1}^{L}(n_{t-1})} \times \\ \times \left(1 - \prod_{k=t}^{T}\left(1 - q_k^L(l_{t-1})\right)\right) \end{array}\right)^{l_{t-1}-1}
\end{aligned}$$

It is easily seen that $u_{t-1}(\alpha_{t-1}, l_{t-1}) < 1$ for all l_{t-1} and $\lim_{l_{t-1}\to\infty} u_{t-1}(\alpha_{t-1}, l_{t-1}) = 1$. Thus, there exist a number $\bar{N}_{t-1}$ such that, for all $n_{t-1} > \bar{N}_{t-1}$, $\arg\max_{l_{t-1}\leq n_{t-1}} u_{t-1} = n_{t-1}$. This implies that $l_{t-1}^{*} = n_{t-1}$. Taking into account that this happens only when $H_{t-1} = 0$, this can be written as $l_{t-1}^{*} = (1 - sign(H_{t-1})) L_{t-1}$. Then the ex-ante pay-off function in stage $t-1$ becomes:

$$\begin{aligned}
\bar{u}_{t-1}(\alpha_{t-1}, n_{t-1}) &= \left(1 - \left(\phi_{t-1}^{L}(\alpha_{t-1}, n_{t-1})\right)^{n_{t-1}}\right) + \\
&+ \left(\phi_{t-1}^{L}(\alpha_{t-1}, n_{t-1})\right)^{n_{t-1}} u_{t-1}\left(\alpha_{t-1}, l_{t-1}^{*}\right) \\
&= 1 - \left(1 - u_{t-1}(\alpha_{t-1}, n_T)\right)\left(\phi_{t-1}^{L}(\alpha_{t-1}, n_{t-1})\right)^{n_{t-1}} \\
&= 1 - (1 - \alpha_{t-1}) \times \\
&\times \left(1 - \alpha_{t-1}\left(1 - \prod_{k=t-1}^{T}\left(1 - q_k^L(n_{t-1})\right)\right)\right)^{n_{t-1}-1}
\end{aligned}$$

Thus, for any $t = 1, ..., T$ there exists $\bar{N}_t$ such that $l_t^{*}(1 - sign(H_t)) L_t$ for all $n_t > \bar{N}_t$. Taking $\bar{N}_t = \bar{N}_1$ ends the proof. ■

Proof of Proposition 3.7.

In what follows, we will use the following notations:

$$\begin{aligned}
\phi_t^L &= (\alpha_t, n_t) \equiv (1 - \alpha_t) q_t^L(n_t) \\
\gamma_t^H(\alpha_t, n_t) &\equiv \frac{\alpha_t}{1 - \phi_t^L(\alpha_t, n_t)}
\end{aligned}$$

First, as it is never optimal to mix high and low signals, $\alpha_t(\alpha_{t-1}, h_{t-1}) \equiv \gamma_{t-1}^{H} sign(h_{t-1})$. We solve the model using backward induction. First, we derive the team's ex-ante pay-off function $\bar{u}_T(\alpha_T, n_T)$ in the last stage, that defines the reduced form pay-off function in stage $T-1$, i.e., $\bar{u}_{T-1}(\alpha_{T-1}, h_{T-1}) = \bar{u}_T(\alpha_T(\alpha_{T-1}, h_{T-1}), h_{T-1})$ for $h_{T-1} > 0$. Maximizing the latter expression w.r.t. h_{T-1} we show that there exists an α_{T-1}^{*} such that $h_{T-1}^{*} = \min(2, H_{T-1})$ and $l_{T-1}^{*} = 1 - sign(H_{T-1})$ for all $\alpha_{T-1} \in \left(\alpha_{T-1}^{*}, 1\right)$.

Next, we derive the ex-ante pay-off function $\bar{u}_{T-1}(\alpha_{T-1}, n_{T-1})$ in stage $T-1$.We generalize it to an arbitrary stage t, i.e., $\bar{u}_t(\alpha_t, n_t)$ using induction arguments, at the same time showing that there exists an $\alpha^*_{t-1} \in (0,1)$ such that the corresponding reduced form pay-off function in stage $t-1$ is maximized at $h^*_{t-1} = \min(2, H_{T-1})$ and $l^*_{t-1} = 1 - sign(H_{t-1})$ for all $\alpha_{t-1} \in (\alpha^*_{t-1}, 1)$.

In stage T, when the sample size is $n_T > 0$ and the prior is $\blacksquare_T$, selector T selects a high signal, if there are, and gets a pay-off $\gamma^H_T(\alpha_T, n_T)$, which happens with probability $1 - (\phi^L_T(\alpha_T, n_T))^{n_T}$. With the remaining probability $(\phi^L_T(\alpha_T, n_T))^{n_T}$ all signals in stage T are low and therefore, he selects one of them and gets zero. Thus

$$\begin{aligned}\bar{u}_T(\alpha_T, n_T) &= \left(1 - (\phi^L_T(\alpha_T, n_T))^{n_T}\right)\gamma^H_T \\ &= \alpha_T \frac{1 - (\phi^L_T(\alpha_T, n_T))^{n_T}}{1 - \phi^T_L(\alpha_T, n_T)}\end{aligned}$$

and, therefore,

$$u_{T-1}(\alpha_{T-1}, h_{T-1}) = \gamma^H_{T-1}\frac{1 - (\phi^L_T)^{h_{T-1}}}{1 - \phi^L_T} = \gamma^H_{T-1}\sum_{k=0}^{h_{T-1}-1}(\phi^L_T)^k$$

Suppose that $H_{T-1} \geq 2$. Then, it is easy to see that $u_{T-1}(\alpha_{T-1}, 2) > u_{T-1}(\alpha_{T-1}, 1)$. On the other hand,

.

$$\begin{aligned}\Delta_{T-1} &\equiv \lim_{\alpha_{T-1}\to 1}\frac{u_{T-1}(\alpha_{T-1}, x) - u_{T-1}(\alpha_{T-1}, x+1)}{1 - \alpha_{T-1}} \\ &= \left(1 - q^L_{T-1}\right)\left(q^L_T(x) - q^L_T(x+1)\right) > 0\end{aligned}$$

Thus, there exists an $\alpha^*_{T-1} \in (0,1)$ such that $h^*_{T-1} = \min(2, H_{T-1})$ for all $\alpha_{T-1} \in (\alpha^*_{T-1}, 1)$. If $H_{T-1} = 0$ player $T-1$ has no better option than taking one low signal and getting zero pay-off. Thus, $l^*_{T-1} = 1 - sign(H_{T-1})$. The ex-ante pay-off in period $T-1$ can now be written as

$$\begin{aligned}\bar{u}_{T-1}(\alpha_{T-1}, n_{T-1}) = \; & u_{T-1}(\alpha_{T-1}, 2) \times \\ & \times\left(1 - (\phi^L_{T-1})^{n_{T-1}} - n_{T-1}(1 - \phi^L_{T-1})(\phi^L_{T-1})^{n_{T-1}-1}\right) + \\ & + u_{T-1}(\alpha_{T-1}, 1)\, n_{T-1}(1 - \phi^L_{T-1})(\phi^L_{T-1})^{n_{T-1}-1} + \\ & + u_{T-1}(\alpha_{T-1}, 0)(\phi^L_{T-1})^{n_{T-1}}\end{aligned}$$

that is,

$$\begin{aligned}\bar{u}_{T-1}\left(\alpha_{T-1},n_{T-1}\right) &= \alpha_{T-1}\sum_{k=0}^{n_{T-1}-1}\left(\phi_{T-1}^{L}\right)^{k}+ \\ &+\alpha_{T-1}\frac{\left(1-\alpha_{T-1}\right)\left(1-q_{T-1}^{L}\left(n_{T-1}\right)\right)}{1-\phi_{T-1}^{L}}\times \\ &\times\left(\sum_{k=0}^{n_{T-1}-1}\left(\phi_{T-1}^{L}\right)^{k}-n_{T-1}\left(\phi_{T-1}^{L}\right)^{n_{T-1}-1}\right)q_{T}^{L}\left(2\right)\end{aligned}$$

Suppose that at stage t the ex-ante pay-off function is given by

$$\begin{aligned}\bar{u}_{t}\left(\alpha_{t},n_{t}\right) &= \alpha_{t}\sum_{k=0}^{n_{t}-1}\left(\phi_{t}^{L}\right)^{k}+\alpha_{t}\frac{\left(1-\alpha_{t}\right)\left(1-q_{t}^{L}\left(n_{t}\right)\right)}{1-\phi_{t}^{L}}\times \\ &\times\left(\sum_{k=0}^{n_{t}-1}\left(\phi_{t}^{L}\right)^{k}-n_{t}\left(\phi_{t}^{L}\right)^{n_{t}-1}\right)\beta_{t}\end{aligned}$$

where $\beta_t \in (0,1)$. Suppose also that there exists an $\alpha_t^* \in (0,1)$such that $h_t^* = \min(2, H_t)$ and $l_t^* = 1 - sign(H_t)$ for all . $\alpha_t \in (\alpha_t^*, 1)$. The corresponding reduced form pay-off function in stage $t-1$ is given by $u_{t-1}(\alpha_{t-1}, h_{t-1}) = \bar{u}_t(\alpha_t, h_{t-1})$. Suppose that $H_{t-1} \geq 2$. Then, it is easy to see that $u_{t-1}(\alpha_{t-1}, 2) > u_{t-1}(\alpha_{t-1}, 1)$. On the other hand,

$$\begin{aligned}\Delta_{t-1} &\equiv \lim_{\alpha_{t-1}\to 1}\frac{u_{t-1}\left(\alpha_{t-1},x\right)-u_{t-1}\left(\alpha_{t-1},x+1\right)}{1-\alpha_{t-1}} \\ &= \left(1-q_{t-1}^{L}\right)\left(q_{t}^{L}\left(x\right)-q_{t}^{L}\left(x+1\right)\right)\left(1-\beta_{t}\right)>0\end{aligned}$$

Thus, there exists an an $\alpha_{t-1}^* > \alpha_t^*$ such that, $\alpha_{t-1} \in \left(\alpha_{t-1}^*, 1\right)$, $h_{t-1}^* = \min(2, H_{t-1})$ and $l_{t-1}^* = 1 - sign(H_{t-1})$ for all $\alpha_{T-1} \in \left(\alpha_{T-1}^*, 1\right)$. The ex-ante pay-off in period $t-1$ can be written as

$$\begin{aligned}\bar{u}_{t-1}\left(\alpha_{t-1},n_{t-1}\right) &= u_{t-1}\left(\alpha_{t-1},2\right)\times \\ &\times\left(1-\left(\phi_{t-1}^{L}\right)^{n_{t-1}}-n_{t-1}\left(1-\phi_{t-1}^{L}\right)\left(\phi_{t-1}^{L}\right)^{n_{t-1}-1}\right)+ \\ &+u_{t-1}\left(\alpha_{t-1},1\right)n_{t-1}\left(1-\phi_{t-1}^{L}\right)\left(\phi_{t-1}^{L}\right)^{n_{t-1}-1}+ \\ &+u_{t-1}\left(\alpha_{t-1},0\right)\left(\phi_{t-1}^{L}\right)^{n_{t}-1} \\ &= \alpha_{t-1}\sum_{k=0}^{n_{t-1}-1}\left(\phi_{t-1}^{L}\right)^{k}+\alpha_{t-1}\frac{\left(1-\alpha_{t-1}\right)\left(1-q_{t-1}^{L}\left(n_{t-1}\right)\right)}{1-\phi_{t-1}^{L}}\times \\ &\times\left(\sum_{k=0}^{n_{t-1}-1}\left(\phi_{t-1}^{L}\right)^{k}-n_{t-1}\left(\phi_{t-1}^{L}\right)^{n_{t-1}-1}\right)\beta_{t-1}\end{aligned}$$

where $\beta_{t-1} = \left(q_t^L(2) + \left(1 - q_t^L(2)\right)\beta_t\right) = 1 - \prod_{k=t}^{T}\left(1 - q_k^L(2)\right)$. Hence, by induction, for any $t = 1, \ldots, T-1$ there exists an $\alpha_t^* \in (0,1)$ such that for all $\alpha_t \in (\alpha_t^*, 1)$ and for all $\tau \geq t$: $h_\tau^* = \min(2, H_\tau)$ and $l_\tau^* = 1 - sign(H_\tau)$. Taking $t = 1$ with $\alpha^* = \alpha_1^*$ ends the proof. ■

Derivation of (2).

For two-stage filtering selection with $q_t^H = 1$ the residual-form pay-off in stage 1 is given by

$$u(\alpha, h_1) = \gamma_1^H \sum_{k=0}^{h_1-1}\left(\left(1-\gamma_1^H\right) q_2^L(h_1)\right)^k$$

The marginal disutility of having an extra high signal, $MU(n) = u(\alpha, n) - u(\alpha, n+1)$ becomes:

$$\begin{aligned} MU(n) &= u(\alpha, n) - u(\alpha, n+1) \\ &= \gamma_1^H \begin{pmatrix} \sum_{k=1}^{n-1}\left(\left(q_2^L(n)\right)^k - \left(q_2^L(n+1)\right)^k\right)\left(1-\gamma_1^H\right)^k - \\ -\left(1-\gamma_1^H\right)^n \left(q_2^L(n+1)\right)^n \end{pmatrix} \\ &> \gamma_1^H\left(1-\gamma_1^H\right)\left(\left(q_2^L(n) - q_2^L(n+1)\right) - (1-\alpha)^{n-1}\left(q_2^L(n+1)\right)^n\right) \end{aligned}$$

Thus, if $\left(q_2^L(n) - q_2^L(n+1)\right) > (1-\alpha)^{n-1}\left(q_2^L(n+1)\right)^n$ for all $n \geq \bar{n}$, then $h_1^* \leq \bar{n}$. ■

4. The optimal ordering of agents in organizations[1]

1 Introduction

As for the previous chapter, the analysis focuses here on collective decision-making. Differently from chapter 3, the only processing limitation I account for here is lack of communication.

Much economic literature has been devoted to analyze the structure and performance of economic organizations. The main motivation behind this branch of research is the consideration that individuals have limited capability in processing information and, consequently, groups of individuals may be able to perform better than any single individual. An economic organization can be viewed as a decision making network in which each node is its constituent agent. The structure of the organization specifies who gathers information, and who communicates what to whom. The basic feature of an organization is that of aggregating the decisions taken by each one of its members separately. Therefore, it is the specific organizational structure that determines the quality of the final decision undertaken by the organization as a whole.

Since it is natural to assume that agents belonging to the same organization differ in ability (some individuals make, on average, more accurate decisions than others) a legitimate question that arises is whether the ordering of heterogenous agents affects the performance of a specific organization. Sah and Stiglitz (1986), and the subsequent literature which follows their approach[2], provide a negative answer to this question. The common feature of this branch of the literature is that of analyzing the performance of different organizational structures in the context of *project evaluation*: several projects, whose quality is unobservable, are assumed to come in streams, and the organization evaluates them sequentially in order to distinguish those that are worth implementing from those that are not. However, there exist many examples in which organizations perform a different task. For instance, the goal of many organizations is that of selecting the *best* out of an arbitrary number of projects (alternatives hereafter) that are simultaneously available. I call this problem, which is different form the one considered by Sah and Stiglitz (1986), *alternatives selection* problem. Consider, for example, the hiring decision of a firm. If the number of vacancies is unlimited

[1] This chapter is based on a joint work with Vladimir Karamychev.

[2] See for example, Hendrikse (1992), Ioannides (1987), Koh (1992a, b, 1993, 1994a, b) and Visser (2000).

then the firm faces an *evaluation* problem as it must distinguish those workers whose marginal productivity is larger than the wage offered form those with a productivity lower than such a wage. If, on the contrary, only one vacancy is available, then the problem is typically a *selection* one as the vacancy should be given to the most productive worker.

In the *project evaluation* framework the terminal payoffs are threefold: the organization gains a profit if it accepts a good alternative, it incurs a loss if it accepts a bad alternative, and gets a payoff of zero if it rejects the alternative. Consequently, the performance of a specific organization is determined by the trade-off between the individuals' errors of not approving good alternatives (Type-I errors) and the errors of approving bad alternatives (Type-II errors). In the *alternatives selection* framework, on the contrary, the terminal payoffs are twofold: if the selected alternative is good the organizations gets a high payoff (say, one), while, if the selected alternative is bad, the payoff is low (say, zero).

In this chapter I focus the attention on a hierarchy because it is the most effective organizational structure to perform the task of selecting the best out of several alternatives. A two-stage selection problem corresponds here to a hierarchical structure where the first agent preselects a subsample of alternatives to be passed to the second stage, and the second agent makes the final selection by choosing one of the preselected alternatives.

A crucial aspect in any organization is the amount of information that can be communicated between agents. It is plausible to think that communication, like decision making, is always imperfect[3]. In order to account for this imperfection I consider two extreme cases, one in which there is perfect communication between agents (denoted by PC), and one in which there is no communication at all (NC). In this model agents are assumed to be fully rational and heterogenous in the accuracy with which they screen the alternatives.

I show that, in contrast to the *evaluation* framework, in the *selection* framework the performance of an organization is generally affected by the ordering of its agents. I find that the performance of the hierarchy improves when the more accurate agent screens first. Moreover, by letting the best agent act first, the organization overcomes the imperfections due to the lack of internal communication and becomes as efficient as a hierarchy in which the information flow is perfect. Not surprisingly, when communication between agents is perfect, the order in which agents are placed does not matter.

It is important to stress that I evaluate the performance of the organization by its gross expected profit, that is, I assume there is no direct cost

[3]For example, communication problems may arise due to the inevitable contamination that occurs in the process of information transmission, or due to a high degree of labour specialization.

involved in processing information. It is indeed the gross expected profit which, in the *project implementation* literature, turns out to be always unaffected by the ordering of heterogenous agents. If one introduces variable evaluation costs, then the ordering of agents matters also in the standard *implementation* framework (see for example Koh (1992a)).

The structure of the chapter is as follows. In section 2 I present the selection problem, in section 3 I investigate the optimal ordering of agents, in section 4 I provide a link between the evaluation approach introduced by Sah and Stiglitz (1986) and the one I take here. Section 5 concludes the chapter.

2 The alternatives selection problem

There is a population of N alternatives. The quality of each alternative, denoted by θ, can be high or low, $\theta \in \{\theta_H, \theta_L\}$. A high quality alternative yields utility of 1, while a low quality alternative yields utility of 0. The share of high-type alternatives in the population is denoted by α. The organization I consider is a hierarchy with two agents denoted by 1 and 2. Agent 1 screens all N alternatives and decides which ones to pass to agent 2. Agent 2 screens the subsample preselected by agent 1 and eventually selects one alternative. The expected quality of the alternative selected by agent 2 determines the payoff of the organization. Each agent in the model is interpreted to be Bayesian, with each of them receiving an imperfect binary signal (high or low) about the quality of each alternative. A signal observed in stage t is denoted by $x_t \in \{h_t, l_t\}$, with $t = 1, 2$ and we assume that

$$\Pr(h_t \mid \theta_H) = \Pr(l_t \mid \theta_L) = q_t \in (1/2, 1) \tag{1}$$

Assumption (1) implies that signals have the *monotone likelihood ratio property* (MLRP), where q_t is the revealing probability of a signal observed in stage t. Notice that q_t denotes also the screening accuracy of agent t. Finally, I assume that all signals in stage 1 and 2 are statistically independent

$$\Pr(x_1, x_2 \mid \theta_i) = \Pr(x_1 \mid \theta_i) \Pr(x_2 \mid \theta_i), i = H, L \tag{2}$$

I model the PC case by assuming that agent 2 observes not only the signal outcomes produced in stage 2 but also those produced in stage 1. By doing so, agent 2 makes his choice based on signals produced in both stages and use the pair (x_1, x_2) to determine the overall likelihood value for each alternative[4]. On the contrary I assume that, in NC case, agent 2 makes his

[4] One can think of agent 1 and agent 2 as two engineers each one evaluating a different feature of several projects. Given the common background, agent 2 can read and understand the report made by agent 1 about each preselected project.

choice based only on signals x_2[5].

The strategies of the two agents are as follows. Agent 1, having observed the signal outcomes of stage 1, passes a number of h_1-alternatives and a number of l_1-alternatives to stage 2. Agent 2 observes the signal outcomes of stage 2 (and of stage 1 in case of PC) and eventually selects the final alternative. The selection rule of the hierarchy, is therefore defined by the strategy profile $S = \{s_1, s_2\}$, where s_1 and s_2 are the strategies sequentially implemented by each one of the two agents.

The exogenous parameters α, q_1 and q_2 induce a distribution over the $(N \times 2)$ matrix of binary signal outcomes, $\mathbf{X} \equiv [x_{jt}] \equiv [\mathbf{x}_1, \mathbf{x}_2]$, where $j = 1, .., N$ is an arbitrary alternative, $t = 1, 2$ is the signalling stage, and $\mathbf{x}_1$ and $\mathbf{x}_2$ are the column vectors of signal realizations in stage 1 and stage 2 respectively. The specific matrix realization $\mathbf{X}$ and the strategy profile S, determine the probability that each alternative j is eventually selected. When alternative j is selected the terminal payoff is $\Pr(\theta_j = \theta_H \mid x_{j1}, x_{j2})$, therefore, the expected payoff of the organization is

$$u = \sum_{\mathbf{X} \in \mathbb{X}} \Pr(\mathbf{X}) \cdot \sum_{j=1}^{N} \Pr(j \text{ is selected} \mid \mathbf{X}) \cdot \Pr(\theta_j = \theta_H \mid x_{j1}, x_{j2}) \qquad (3)$$

where $\mathbb{X}$ is the set of possible matrix realizations. The last probability in (3) is a standard Bayesian update of the prior. The probability in the middle of (3) is, on the other hand, a function of the strategy profile S and the information environment. Then, in the PC case

$$\begin{aligned} \Pr(j \text{ is selected} \mid \mathbf{X}) &= \Pr(j \text{ is selected in stage 1} \mid \mathbf{x}_1) \times \\ &\quad \times \Pr(j \text{ is selected in stage 2} \mid \mathbf{x}_1, \mathbf{x}_2) \end{aligned}$$

while in the NC case

$$\begin{aligned} \Pr(j \text{ is selected} \mid \mathbf{X}) &= \Pr(j \text{ is selected in stage 1} \mid \mathbf{x}_1) \times \\ &\quad \times \Pr(j \text{ is selected in stage 2} \mid \mathbf{x}_2) \end{aligned}$$

Finally, I assume that agents are fully rational and that the distribution of types and the structure of the organization are common knowledge.

3 The optimal ordering of agents

In this section I investigate whether the ordering of heterogenous agents affects the performance of the organization. I model heterogeneity by assuming that one agent has a high screening accuracy, q^H while the other has

[5] Think of agent 1 is an engineer and agent 2 as a salesman. The specialization of labour makes it impossible for the salesman to understand the technical reports written by the engineer.

a low screening accuracy, $q^L < q^H$. I denote by $u\left(q^H, q^L\right)$ $\left(u\left(q^L, q^H\right)\right)$ the expected payoff of the hierarchy when the agent with better accuracy acts first (second).

I start from the benchmark PC case and, not surprisingly, I find that the performance of the hierarchy is not affected by the ordering of its agents.

Proposition 4.1. $u^{PC}\left(q^H, q^L\right) = u^{PC}\left(q^L, q^H\right).$

Proof. *The perfect communication environment and the rationality of agents ensure that passing all alternatives to the second stage is always optimal, and that the optimal selection rule S can be represented by a correct preference relation over all possible signal pairs. If $q_1 = q^H > q^L = q_2$ the preference relation is*

$$(h_1, h_2) \succ (h_1, l_2) \succ (l_1, h_2) \succ (l_1, l_2) \tag{$P1$}$$

If, on the contrary $q_1 = q^L < q^H = q_2$, the preference relation is

$$(h_1, h_2) \succ (l_1, h_2) \succ (h_1, l_2) \succ (l_1, l_2) \tag{$P2$}$$

Such preferences ensure that the hierarchy ranks correctly the rows (alternatives) of the signal realizations' matrix, and eventually selects the one with the higher overall likelihood of being of high type. Since $\Pr(\mathbf{X}) = \Pi_{j=1}^{N} \Pr(x_{1j}, x_{2j})$, *it follows form Bayes' rule that the probability of observing the matrix $[\mathbf{x}_1, \mathbf{x}_2]$, conditional on $q_1 = q^H$ and $q_2 = q^L$, equals the probability of observing the matrix $[\mathbf{x}_2, \mathbf{x}_1]$, conditional on $q_1 = q^L$ and $q_2 = q^H$. That is, given an arbitrary matrix, inverting the order of q^H and q^L induces an identical probability over the matrix in which the order of the columns is also inverted. Such symmetry, together with the symmetry of preferences $P1$ and $P2$, and the fact that the terminal payoff* $\Pr(\theta_H \mid x_1, x_2)$ *does not depend on the order in which q^H and q^L are placed, implies that, indeed, $u^{PC}\left(q^H, q^L\right) = u^{PC}\left(q^L, q^H\right)$.* ■

I now consider a hierarchy in which communication between agents is not possible. I find that, here, the ordering of agents affects the performance of the organization: by having the best screener in the first place the hierarchy achieves a higher payoff. Moreover, by letting the best agent act first, the hierarchy overcomes the imperfections due to the lack of internal communication and turns out to be as efficient as a hierarchy in which the information flow is perfect.

Proposition 4.2. $u^{NC}\left(q^L, q^H\right) < u^{NC}\left(q^H, q^L\right) = u^{PC}\left(q^H, q^L\right).$

Proof. *If $q_1 = q^H > q^L = q_2$ the preference relation over all possible signal pairs is*

$$(h_1, h_2) \succ (h_1, l_2) \succ (l_1, h_2) \succ (l_1, l_2) \tag{$P1$}$$

It is clear from $(P1)$ *that selecting only* h_1*-alternatives is an optimal strategy for agent 1. The reason is that, independently of the signal realizations of stage 2,* l_1*-alternatives will never produce an overall likelihood value larger than that of* h_1*-alternatives. Since it is optimal for agent 2 to select a* h_2*-alternative, the optimal strategy profile ensures indeed that the alternative eventually selected is the one with the higher likelihood value. This proves* $u^{NC}\left(q^H, q^L\right) = u^{PC}\left(q^H, q^L\right)$. *If, on the contrary,* $q_1 = q^L < q^H = q_2$, *the preference relation is*

$$(h_1, h_2) \succ (l_1, h_2) \succ (h_1, l_2) \succ (l_1, l_2) \qquad (P2)$$

Assume first that agent 1 passes only h_1*-alternatives. With positive probability all preselected alternatives produce signals* l_2 *and at least one* l_1*-alternative produces a signal* h_2*. When this event occurs the payoff-dominant pair* (l_1, h_2) *cannot be selected because it has been previously discarded by agent 1. Assume now that some (possibly all)* l_1*-alternatives are passed in addition to* h_1*-alternatives . There exists a positive probability that all preselected alternatives generate identical signals in stage 2 making it impossible for agent 2 to select the payoff-dominant alternative with probability 1. Thus, for any strategy adopted by agent 1, there is always a chance that the alternative with the higher overall likelihood value is not selected. This proves* $u^{NC}\left(q^L, q^H\right) < u^{NC}\left(q^H, q^L\right)$. ■

The intuition behind proposition 4.2 is clear. If the more accurate screener acts first, he does not need to pass to the second stage alternatives that produced a low signal because, independently of the signal realizations of stage 2, they will never produce a likelihood value larger than that of alternatives that produced a high signal in stage 1. Therefore agent 2 knows that he always receives a subsample of projects with identical x_1-signal outcomes. Consequently, he does not need any explicit information about the signal outcomes of stage 1 in order to make the right choice: by selecting an alternative that produced a high signal in stage 2, agent 2 automatically selects the best (in expected terms) alternative.

Such argument can be generalized to an arbitrary number of stages, though, a more stringent requirement over the relative differences in agents' accuracy is needed. If there are T selection stages, ordering the agents in terms of decreasing accuracy $q_1 ... < q_t < ... < q_T$, solves the problems due to imperfect communication only if, at any stage t, it is optimal for agent t to pass only h_t-alternatives. Such strategy is indeed optimal provided that q_t is such that $\Pr(\theta_H \mid h_t, l_{t+1}, .., l_T) \geq \Pr(\theta_H \mid l_t, h_{t+1}, .., h_T)$, for any $t = 1, .., T$.

4 Evaluation versus selection

I show here how the traditional project evaluation problem can be restated in terms of a selection model. Consider the problem of deciding whether or not to implement a project of unknown quality, θ. An organization facing such a problem is actually undecided between two alternatives: alternative 1 stands for *"implement the project"* while alternative 2 stands for *"do not implement the project"*. Therefore an evaluation problem can be restated as a selection problem between two alternatives, provided we impose the following restrictions on signal outcomes

$$\Pr(h_{1t} \mid \theta_H) = \Pr(l_{1t} \mid \theta_L) = q_t \in (1/2, 1), \ t = 1, 2 \tag{4}$$

$$\Pr(h_{1t} \mid l_{2t}) = \Pr(h_{1t} \mid l_{2t}) = 1, \ t = 1, 2 \tag{5}$$

Condition (4) means that, it the project is good (bad), then each screening stage is more likely to suggest in favor of implementing (not implementing) the project. Moreover, (5) means that signal outcomes across alternatives are perfectly correlated: if the report of a screener is in favor of implementing the project it is necessarily against not implementing it, and vice versa. Notice that this new formulation is a more restricted[6] model than the one presented in section 1 as, here, I do not allow for the possibility that two alternatives generate identical signals. This also means that $1 - q_t$ is the probability that screener t is inclined to accept (reject) a bad (good) project and the standard analysis of the trade off between Type-I and Type-II errors can be applied.

5 Conclusion

In this section I stress in more details the differences between the assumptions we use in the current chapter and those present in the existing literature. In most of the literature (see for example, Hendrikse (1992), Ioannides (1987), Koh (1992a, b, 1993, 1994a, b) Sah and Stiglitz (1986)) it is assumed that agents are not rational. They are characterized by a pair of probabilities with which they accept good and bad alternatives. These probabilities are determined exogenously and do not reflect the organizational structure, nor the positions in which agents are placed. Visser (2000), instead, comes close to my approach as he also assumes fully rational heterogeneous individuals that cannot communicate the information they posses. However, in Visser (2000) the ordering of agents does not affect the expected payoff of the organization (proposition 3 and proposition 5 in his paper). The intuition behind this result is the symmetry of the organizational decision

[6] It is more restricted because we impose perfect correlation between signals of two different alternatives: if one is high the other is necessarily low, and vice versa.

problem and of the knowledge agents have about the structure and each other's rationality. The novel *selection* framework I use here is responsible for the different results I obtain. Individual rationality itself is not sufficient to preserve the symmetry of the organizational decision problem. For this to be the case communication must be perfect, otherwise such symmetry fails to hold. Our analysis also provides a solution to this problem as I show that the lack of communication can be overcome by an appropriate ordering of the agents.

Finally, one may argue that, in real life, there exist also examples in which the more accurate screening stages are placed at higher hierarchical levels. My findings seems to address that such practice is not optimal. However, it is important to stress that I didn't account for the direct costs usually involved in any information-processing phase. As it is natural to assume that the cost of processing information increases with the accuracy with which information is processed, it might be the case that having higher accuracy at higher stages is optimal for the simple reason that, at later stages, less information is processed. Whenever such ordering is in place, the contribution of our paper is that of identifying the indirect cost that might emerge due to a lack of internal communication.

5. Individual decision making and imperfect information processing[1]

1 Introduction

After the last two chapters in which I have highlighted the impact of imperfect information processing for a group of individuals, I shift the attention here to individual decision-making. The main interpretation of imperfect information processing in this chapter is that of being a cognitive limitation that prevents a decision-maker from choosing the best feasible alternative.

One of the fundamental results of Choice Theory is the equivalence between maximization of an objective function and the Weak Axiom of Revealed Preferences (WARP). This paper develops a model in which the choice function of a maximizing decision-maker violates WARP.

In accordance with Choice Theory, any choice function assigns a utility maximizing alternative to any given set of alternatives. Choice Theory does not specify how a decision-maker chooses a utility maximizing alternative. Implicitly, it assumes that the decision-maker evaluates his objective function for each alternative, and then chooses an alternative with the largest value. In other words, it assumes that the decision-maker uses a process of eliminating all but one alternative, (which I call a selection procedure) which yields a utility maximizing alternative with certainty.

However, sometimes this assumption is too demanding. For example, evaluation of the expected utility function requires exact knowledge of prior probabilities and correct beliefs about actions taken by others. Similarly, evaluation of the indirect (reduced form) utility function is not possible without knowledge of all potential consequences of each alternative on the direct utility function. Such computational requirements can be far beyond humans' abilities due to cognitive limitations or the complexity of the environment (Simon, 1955). In this chapter, I consider a decision-maker who is unable to evaluate his objective function for all given alternatives.

Due to this inability, the decision-maker has to use another selection procedure, which does not require evaluations of his objective function. There are many such selection procedures. However, none of these procedures can guarantee that the chosen alternative maximizes the objective function. In order to choose an alternative, the decision-maker I model here chooses a selection procedure, and, second, he uses this procedure for choosing an alternative.

[1]This chapter is based on a joint work with Vladimir Karamychev and Peran Van Reeven.

The value of the objective function (value hereafter) that this chosen alternative yields is a random variable, and its distribution is determined by the selection procedure. Hence, choosing a selection procedure is equivalent to choosing this value distribution. Different criteria can be used in order to compare value distributions. In the current chapter, I assume that the decision-maker chooses the procedure that minimizes the type-II error, i.e., one minus the probability that the chosen alternative does not maximize the objective function. In other words, I characterize each value distribution and, therefore, each selection procedure, by the probability of choosing an alternative with the highest value. As a result, all alternatives come into two types: alternatives with the highest value, which I call high-type alternatives, and all the other alternatives, which I call low-type alternatives. I take the probability of selecting a high-type alternative as the objective function of the decision-maker for choosing a selection procedure. When the procedure is chosen, the choice function is fully determined by that procedure.

My task is two-fold. First, I want to identify the selection procedure that minimizes the type-II error. Second, I investigate the properties of the choice function induced by this procedure. In order to do so, I analyze the Choice-By-Elimination model of Tversky (1972), called CBE-model hereafter. In this model, each alternative is represented by a vector of aspects. Aspects can be considered as good-characteristics of Lancaster (1966) and provide partial information about the type of each alternative in the spirit of Milgrom (1981). If an alternative has a certain aspect then it is good news about its value. If, on the other hand, an alternative does not have this aspect, then it is bad news. The elimination process takes place in stages. Each aspect corresponds to a stage in which the decision-maker decides which alternatives he eliminates and which alternatives he selects and passes to the next stage. Aspects are used sequentially and the selection procedure continues until only one alternative remains.

I get the following results. First, there exists a unique selection procedure which minimizes the type-II error. It eliminates an alternative at a stage if and only if it is the only alternative without the aspect. In accordance with its properties, I call this procedure 'Single-Worst-Elimination', SWE hereafter. SWE dominates all other procedures irrespective of the prior distribution of types in the initial set of alternatives, of the size of this initial set, and of the informational accuracy of aspects. In two special limiting cases, when informational accuracy is high, 'Elimination-By-Aspects' (Tversky, 1972), EBA hereafter, and 'Satisficing' (Simon, 1955), SAT hereafter, become optimal selection procedures. Second, the choices induced by SWE are always transitive, but may violate WARP.

The rest of the chapter is organized as follows. Section 2 states the model which is then analyzed in section 3. Section 4 discusses the optimality of SAT and EBA in special limiting cases of the model. Section 5 analyses the choice function induced by SWE. Section 6 discusses modeling assumptions and concludes the paper. The appendix contains a proof of the first proposition.

2 The model

Consider a decision-maker who has choose a single alternative from a given set $\mathbf{X} = \{x_1, .., x_N\} \subset \mathbf{A}$, where $\mathbf{A}$ is a grand set of alternatives. Each alternative is either a high or low type so that $\mathbf{A}$ is a union of two disjoint sets of high and low type alternatives, $\mathbf{A} = \mathbf{\Theta}_H \cup \mathbf{\Theta}_L$and $\mathbf{\Theta}_H \cap \mathbf{\Theta}_L = \varnothing$. For a decision-maker, an alternative is an infinite vector of binary aspects, $x_i = \left(a_i^1, .., a_i^t, ...\right)$, where $a_i = \{0, 1\}$. If $a_i^t = 1$, we say that alternative i has aspect t whereas if $a_i^t = 0$, we say that alternative i does not have aspect t.

Aspects provide imprecise information about the true type of x_i. In particular, we assume that all aspects are equally informative, that a high type alternative has an aspect t with a probability q_H, and that a low type alternative has that aspect with a probability q_L:

$$q_H \equiv \Pr\left(a_i^t = 1 \mid x_i \in \mathbf{\Theta}_H\right) > q_L \equiv \Pr\left(a_i^t = 1 \mid x_i \in \mathbf{\Theta}_L\right), t = 1, ..., \infty$$

A selection procedure takes place in stages. We denote the set of alternatives at the beginning of stage t by $\mathbf{X}_t \subseteq \mathbf{X}$, $\mathbf{X}_t \neq \varnothing$. The decision-maker uses one aspect in each selection stage and, hence, in stage t he observes a set of aspects $\mathbf{a}^t = \left\{a_i^t \mid x_i \in \mathbf{X}_t\right\} = \varnothing$. Having observed this set of aspects $\mathbf{a}^t$, the decision-maker selects a subset of alternatives $\mathbf{X}_{t+1} \subseteq \mathbf{X}_t$ and passes it to the next stage. All alternatives $x_i \in \mathbf{X}_t \setminus \mathbf{X}_{t+1}$are eliminated in stage t. Hence, any selection procedure S is characterized in stage t by the correspondence $\mathbf{X}_{t+1} = \mathbf{X}_{t+1}\left(\mathbf{a}^t\right) \equiv \mathbf{S}_t\left(\mathbf{X}_t\right)$, and is fully characterized by a set of such correspondences for all stages $t = 1, ..., \infty$, so that we write $S = \{\mathbf{S}_t\}$.

Starting from $t = 1$ and $\mathbf{X}_1 = \mathbf{X}$, the decision-maker sequentially apply correspondences $\mathbf{S}_t$ so that $\mathbf{X}_{t+1}$ can be written as a superposition of $\mathbf{S}_t$ as follows:

$$\mathbf{X}_{t+1} = \left(\mathbf{S}_t \circ ... \circ \mathbf{S}_1\right)(\mathbf{X}) = \mathbf{S}^{(t)}(\mathbf{X})$$

One important property of any selection procedure S is that $\mathbf{S}_t(\{x_k\}) = \{x_k\}$ for all t and k, i.e., if $\mathbf{X}_t$ consists of a single alternative x_k, S necessarily passes this alternative to the next stage. This allows us to define a choice function of an individual who follows a selection procedure S as follows:

$$C(\mathbf{X}, S) = \lim_{t \to \infty} \mathbf{S}^{(t)}(\mathbf{X})$$

i.e., $C(\mathbf{X}, S)$ is the alternative that a decision-maker, who follows a selection procedure S, chooses from the set $\mathbf{X}$. The limit here always exists as $\mathbf{X}$ is finite and $\mathbf{S}^{(t+1)}(\mathbf{X}) \subseteq \mathbf{S}^{(t)}(\mathbf{X}) \subseteq \mathbf{X}$ for all $\mathbf{X} \subset \mathbf{A}$. For some selection procedures, however, $C(\mathbf{X}, S)$ is not a singleton. For example, if the selection procedure is such that $\mathbf{S}^{(t)}(\mathbf{X}_t) = \mathbf{X}_t$, then $\mathbf{X}_t = \mathbf{X}$ for all t, and

$C(\mathbf{X},S) = \mathbf{X}$. We exclude such selection procedures from the analysis and only consider procedures that choose a single alternative.

The chosen alternative can either be a high type or a low type alternative. The decision-maker chooses a selection procedure S^*, which maximizes the probability that the chosen alternative is of high type:

$$S^* = \arg\max_S \Pr\left(C(\mathbf{X},S) \subset \mathbf{\Theta}_H\right)$$

If it exists, we call this procedure 'optimal selection procedure'. Our objective is to identify S^* and explore properties of the induced choice function $C(\mathbf{X},S^*)$.

3 Analysis

We denote by N_t the number of alternatives at the beginning of stage t, i.e., $N_t = \#X_t$, where $\#$ counts the number of elements in the set that follow. The set of aspects $\mathbf{a}^t$ that the decision-maker observes at stage t can be written as follows:

$$\mathbf{a}^t = \left\{a_i^t \mid x_i \in \mathbf{X}_t, a_i^t = 1\right\} \cup \left\{a_i^t \mid x_i \in \mathbf{X}_t, a_i^t = 0\right\}$$

The first term denotes alternatives that have the aspect, which we call 'good' alternatives. The second term denotes alternatives that do not have the aspect, which we call 'bad' alternatives. Using the following notation for the numbers of good and bad alternatives respectively,

$$G_t \equiv \#\left\{a_i^t \mid x_i \in \mathbf{X}_t, a_i^t = 1\right\} \text{ and } B_t \equiv \#\left\{a_i^t \mid x_i \in \mathbf{X}_t, a_i^t = 0\right\}$$

we write $N_t = G_t + B_t$.

The decision-maker does not distinguish between good alternatives. Therefore, a selection procedure S in stage t, i.e., the correspondence $\mathbf{S}_t(\mathbf{X}_t)$, only specifies the number of good alternatives g_t to be passed to the next stage; it cannot specify which of those good alternatives are passed. Similarly, g_t depends only on the number of good alternatives observed, G_t, and on the number of bad alternatives observed B_t; it cannot depend on which alternatives are good and which alternatives are bad. By the very same reason, $\mathbf{S}_t(\mathbf{X}_t)$ specifies only the number of bad alternatives b_t to be selected, where b_t is a function of G_t and B_t. Hence, a selection procedure S is fully characterized by the pair of functions $(g(t,G,B), b(t,G,B))$ with $g(t,G,B) + b(t,G,B) \geq 1$, $0 \leq g(t,G,B) \leq G$ and $0 \leq b(t,G,B) \leq B$ so that $g_t = g(t,G,B)$ and $b_t = b(t,G,B)$. For brevity, we denote this selection procedure as $S(g,b)$. An optimal selection procedure, if it exists, will be denoted by $S^* = S(g^*,b^*)$

The following examples demonstrate how several known selection procedures can be written in terms of (g,b)-functions.

Example 1 *Satisficing (Simon, 1955). Let* $(g,b) = (1,0)$ *if* $G \geq 1$*, and* $(g,b) = (0,1)$ *if* $G = 0$*. This selection procedure randomly selects one of the first alternatives in the set that has an aspect. We denote this selection procedure by* S^{SAT}.

Example 2 *Elimination-By-Aspects, EBA hereafter (Tversky, 1972). Let* $(g,b) = (G,0)$ *if* $G \geq 1$*, and* $(g,b) = (0,1)$ *if* $G = 0$*. This selection procedure selects only alternatives that have an aspect. If no alternative has an aspect, one of them is randomly selected. We denote this selection procedure by* S^{EBA}.

Example 3 *Choice of a random alternative. Let* $(g,b) = (1,0)$ *if* $B = 0$*, and* $(g,b) = (0,1)$ *if* $G = 0$ *and* $(g,b) = (G,B)$ *if* $G > 0$ *and* $B > 0$*. This selection procedure randomly selects one alternative independently on the aspects that it has.*

The probability that the finally chosen alternative is of high type, that is $\Pr(C(\mathbf{X},S) \subset \mathbf{\Theta}_H)$, depends on the selection procedure S, on the total number alternatives in the set $N = \#\mathbf{X}$, and on the number of high type alternatives $N = \#(\mathbf{X} \cap \mathbf{\Theta}_H)$. Hence, it defines a function

$$U(S,N,N_H) = \Pr(C(\mathbf{X},S) \subset \mathbf{\Theta}_H)$$

,
which is the objective function of the decision-maker, who maximizes $U(S,N,N_H)$ over all feasible selection procedures. By feasibility we refer to $\#C(\mathbf{X},S) = 1$ for all for all $\mathbf{X} \subset \mathbf{A}$. In this notation, the problem of a decision-maker is as follows:

$$\begin{aligned} S^* &= \arg\max_S U(S, \#\mathbf{X}, \#(\mathbf{X} \cap \mathbf{\Theta}_H)) \\ s.t &: \#C(\mathbf{X},S) = 1 \text{ for all } \mathbf{X} \subset \mathbf{A} \end{aligned}$$

In the following proposition, some properties of optimal selection procedures are derived.

Proposition 5.1 *For any* $q_H \in (0,1)$*,* $q_L \in (0,1)$*,* $N = 1,..,\infty$*,* $N_H = 1,...,N-1$*, an optimal selection procedure* $S(g^*,b^*)$ *must satisfy the following properties:*

a) $g^*(t,G,B) = G$*, i.e., all good alternatives are always selected;*
b) Either $b^*(t,G,B) = 0$ *or* $b^*(t,G,B) = B$*, i.e., either all bad alternatives or none of them are selected;*
c) $b^*(t,G,B) = B$ *for* $B \geq 2$*, i.e., if there are at least two bad alternatives, then all bad alternatives are selected.*
d) The maximum $U^*(N,N_H)$ *strictly increases in* N_H.

The proof of Proposition 5.1 is in the appendix and is made by induction. If $N = 1$, the only feasible selection procedure is $(g, b) = (G, B)$, which is trivially optimal and satisfies the properties (a)-(c) of Proposition 5.1. Moreover, the probability of choosing a high type alternative is formally $U(S, 1, N_H) = N_H$, so that $U^*\left(1, N^H\right) \equiv N_H$ strictly increases in $N_{H.}$

Suppose that for some number $K \geq 1$, procedure $S(g^*, b^*)$ is optimal for all $N \leq K$, i.e., it satisfies the properties (a)-(d) of Proposition 5.1. Then, suppose that at a selection stage t, the number of alternatives is . $N_t = K+1$. Suppose also that the decision-maker observes G_t good alternatives and $B_t = N_t - G_t$ bad alternatives, and selects g_t and b_t of them correspondingly so that $g_t + b_t \leq N_t - 1 = k$. In this case, for any fixed prior distribution of the number of high type alternatives, the posterior distribution in this stage and, hence, the prior distribution of N_{t+1}^H in stage $t+1$, is determined by (g_t, b_t). Moreover, from stage $t+1$ on, the number of alternative does not exceed K so that the procedure $S(g^*, b^*)$ will be used, and the conditional on N_{t+1}^H probability of choosing a high type alternative will be $U^*\left(N_{t+1}, N_{t+1}^H\right)$. By the induction assumption, $U^*\left(N_{t+1}, N_{t+1}^H\right)$ increases in N_{t+1}^H. Thus, if the posterior distribution induced by (g_t', b_t') first-order stochastically dominates the posterior distribution induced by (g_t, b_t), then selecting (g_t, b_t) in stage t is not optimal.

Using this first-order stochastic dominance criterion, we compare different selection procedures without evaluating $U(S, N, N_H)$. First, we show that if $g_t < G_t$, i.e., not all good alternatives are selected, then selecting $(g_t + 1, b_t)$, i.e., one good alternative more, and eliminating a single bad alternative later is strictly better. This proves part (a) of Proposition 5.1. Second, we show that if $b_t < B_t$, i.e., not all bad alternatives are selected, then selecting $(g_t^*, b_t + 1)$, i.e., one bad alternative more, and eliminating a single bad alternative later is weakly better. It is strictly better if $b_t \geq 1$. This proves parts (b) and (c) of Proposition 5.1. Lastly, because the eliminated alternative is a bad alternative, i.e., an alternative without an aspect, and due to the single-crossing property $q_H > q_L$, the highest probability of choosing a high type alternative increases in the number of high type alternatives in the initial set. This proves part (d) of Proposition 5.1.

The following properties of CBE-model in general and of the optimal selection procedure in particular are worth mentioning here. First, it is not true that having more high type alternatives in the initial sample increases chances of a decision-maker to choose one of them. The following example demonstrates this point.

Example 4 *Let us take $N = 3$, $q_L = 0.5$ and the selection procedure such that $(g, b) = (1, 0)$ if $(G_t, B_t) = (1, 2)$, $(g, b) = (0, 1)$ if $(G_t, B_t) = (2, 1)$, and $(g, b) = (G_t, B_t)$ otherwise. In other words, a decision-maker who follows selects either a unique good alternative or a unique bad alternative in each selection stage, and selects all three alternatives otherwise. It is easy to get* *the following expressions for the objective function* $U\left(\bar{S},3,N_H\right)$*:*

$$\begin{aligned} U\left(\bar{S},3,0\right) &= 0 \\ U\left(\bar{S},3,1\right) &= \frac{1}{3} \\ U\left(\bar{S},3,2\right) &= \frac{4q_H\left(1-q_H\right)}{2-\left(q_H\right)^2-\left(1-q_H\right)^2} \\ U\left(\bar{S},3,3\right) &= 1 \end{aligned}$$

It follows that $U\left(\bar{S},3,2\right) < U\left(\bar{S},3,1\right)$ *if* $q_H\left(1-q_H\right) < 0.1$*, so that the decision-maker is better-off when the initial sample has one high type alternative rather than two high type alternatives.*

This example stresses the fact that it is not the function $U\left(S,N,N_H\right)$ to be monotonically increasing in N_H, but its maximum across all possible selected procedures, i.e., $U^*\left(N,N_H\right)$.

Second, it is not generally true that the decision-maker must select all good alternatives at all stages. Let us go back to Example 4. Suppose that q_H is very close to one, and all three alternatives have the first aspect. Then, if the decision-maker selects all three alternatives, he gets $U\left(\bar{S},3,2\right)$ in stage two, i.e., almost zero. Obviously, selecting all good alternatives in stage one is not optimal if the decision-maker follows $\bar{S}$ later on. Hence, the optimal number of good alternatives selected depends on which procedure the decision-maker uses in later stages.

There are two selection procedures that satisfy the properties (a)-(c) of Proposition 5.1. The first procedure is the 'Always Pass' procedure with $(g,b)=(G,B)$, i.e., all alternatives, whether they are good or bad, are passed to the next selection stage. Obviously, this procedure is not feasible. The other selection procedure is such that all alternatives must be selected in each stage, unless only one alternative in the sample is bad. In this case, all good alternatives must be selected, and the single bad alternative must be eliminated. Formally:

$$\left(g^*,b^*\right)=\begin{cases}(G,0) & \text{if } G\geq 0 \text{ and } B=1\\ (G,B) & \text{otherwise}\end{cases}$$

In accordance with its properties, we call this selection procedure 'Single Worst Elimination', abbreviated as SWE and denoted as S^{SWE}. SWE prescribes only to eliminate an alternative if it is the single worst (without the aspect) alternative at a certain selection stage. Thus, we have proven the following result.

Proposition 5.2. *For any* $q_H\in(0,1)$*,* $q_L\in(0,1)$*,* $N=1,..,\infty$*, SWE is an optimal selection procedure. For* $N_H=1,...,N-1$*, SWE is a unique optimal selection procedure.*

Let us compare SWE with SAT and EBA. In accordance with SAT, a decision-maker selects one good alternative. SWE is a mirror image of SAT. In accordance with SWE, a decision-maker eliminates one bad alternative. Similar to SWE, EBA selects all good alternatives. The difference is that EBA eliminates all bad alternatives whereas SWE only eliminates a bad alternative if it is a unique bad alternative at a stage. Hence, SWE uses more selection stages than EBA, and, consequently, results in more accurate choices than EBA.

The number of selection stages at which alternatives are eliminated allows for another characterization of selection procedures. Each selection procedure guarantees that the chosen alternative has at least a certain number of aspects. SWE maximizes this minimum number of aspects that the chosen alternative has. SWE chooses an alternative with at least $(N-1)$ aspects. All the other $(N-1)$ alternatives were eliminated in $(N-1)$ selection stages in which the chosen alternative had the corresponding aspects. In contrast, SAT and EBA guarantee that the chosen alternative has far fewer aspects. SAT chooses an alternative that has at least one aspect, because this alternative is the very first alternative with an aspect. EBA might choose an alternative that does not have aspects at all, because it requires that one alternative must be selected even if all alternatives in the first stage do not have the first aspect.

However, SWE is not the only procedure that satisfies this 'max-min'-property. For example, a procedure which specifies that in each stage a single alternative must be eliminated if, and only if, all alternatives have the aspect, i.e., $(g,b) = (G-1,0)$ if $B = 0$ and $G \geq 2$, and $(g,b) = (G,B)$ otherwise, also guarantees that the chosen alternative has at least $(N-1)$ aspects. Since such a procedure chooses an alternative at random, it is clearly not optimal. Hence, the 'max min'-property is a necessary but not a sufficient condition for the optimality of a selection procedure.

4 Optimality of SAT and EBA

In section 4, we have shown that SWE is a unique optimal selection procedure for $0 < q_L < q_H < 1$. In this section, we show that if low type alternatives never have aspects, i.e., if $q_L = 0$, then SAT is optimal. Alternatively, if high type alternatives always have aspects, i.e., if $q_H = 1$, then EBA is optimal.

The main implication of the assumptions $q_L = 0$ or $q_H = 1$ is that SWE is not feasible any more. Indeed, if $q_L = 0$ and the initial sample of alternatives consists of more than two low type alternatives, then there will be at least two bad alternatives in each selection stage. Hence, SWE does not result in a choice of a single alternative and, therefore, is not feasible. The same problem occurs if $q_H = 1$ and there are at least two high type alternatives in the sample. At the same time, all other selection procedures are strictly dominated and, therefore, not optimal by the very same reasoning as in the proof of Proposition 5.1. Thus, with infinitely many aspects, an optimal selection procedure does not exist.

In order to restore the existence of an optimal selection procedure, we modify the model as follows. We assume that the number of aspects is finite and equals to T. Then, if more than one alternative is selected in stage T, one of these alternatives will be chosen at random at stage $T+1$.

If $q_L = 0$, a low type alternative never has an aspect so that any good alternative is definitely a high type alternative. Selecting any number of good alternatives followed by any selection procedure afterwards is optimal. If there are no good alternatives in the sample, it is optimal to select all alternatives. Hence, if $q_L = 0$, there are multiple optimal selection procedures:

$$S^{SAT} = (g^*, b^*) = \begin{cases} (\tilde{g}, 0) & \text{if } G \geq 1 \\ (0, B) & \text{if } G = 0 \end{cases}$$

for any arbitrary $\tilde{g} = 1, .., \tilde{G}$ which may depend on t.

All these selection procedures are essentially SAT because they wait for good alternatives and randomly select one of them.

If $q_H = 1$, a high type alternative always has an aspect so that any bad alternative is definitely a low type alternative. In the presence of one or more good alternatives, removing all bad alternatives is optimal. If there are no good alternatives, then all alternatives are of low type, and all selection procedures are pay-off equivalent. Hence, if $q_H = 1$, there are multiple optimal selection procedures:

$$S^{EBA} = (g^*, b^*) = \begin{cases} (G, 0) & \text{if } G \geq 1 \\ \left(0, \tilde{b}\right) & \text{if } G = 0 \end{cases}$$

for an arbitrary $\tilde{b} = 1, .., \tilde{B}$, which may depend on t.

All these selection procedures are essentially EBA because, if feasible, they eliminate in every stage all bad alternatives.

5 Choice Function of SWE

In this section, we analyze the choice function and show that this choice function violates WARP.

Proposition 5.3. *The choice function $C\left(\mathbf{X}, S^{SWE}\right)$ is transitive but does not respect WARP.*

Proof. Let be $x_1, x_2, x_3 \in \mathbf{A}$ any three alternatives such that

$$C\left(\{x_1, x_2\}, S^{SWE}\right) = x_1 \text{ and } C\left(\{x_2, x_3\}, S^{SWE}\right) = x_2$$

We define $t_1 = \min\left\{t \mid a_1^t = 1, a_2^t = 0\right\}$, i.e., t_1 is the selection stage in which SWE eliminates alternative x_2 and keeps alternative x_1 in making a choice from the set $\{x_1, x_2\}$. Similarly, we define $t_2 = \min\left\{t \mid a_2^t = 1, a_3^t = 0\right\}$. By construction, $t_1 \neq t_2$. If $t_1 < t_2$, then all three alternatives have equal aspects, i.e., $a_1^t = a_2^t = a_3^t$, for $t = 1, .., t_1 - 1$. By construction, $a_1^{t_1} = 1$ and $a_2^{t_1} = 0$ and . Since $C\left(\{x_2, x_3\}, S^{SWE}\right) = x_2$ and x_3 is eliminated at stage $t_2 > t_1$, it must be that $a_3^{t_1} = a_2^{t_1} = 0$. Hence, $C\left(\{x_1, x_3\}, S^{SWE}\right) = x_1$ because x_3 is the single worst alternative at stage t_1. Alternatively, if $t_2 < t_1$, then all three alternatives have equal aspects, i.e., $a_1^t = a_2^t = a_3^t$, for $t = 1, .., t_2 - 1$. By construction, $a_2^{t_2} = 1$ and $a_3^{t_1} = 0$. Since $C\left(\{x_1, x_2\}, S^{SWE}\right) = x_1$ and x_2 is eliminated at stage $t_1 > t_2$, it must be that $a_1^{t_1} = a_2^{t_1} = 1$. Hence, $C\left(\{x_1, x_3\}, S^{SWE}\right) = x_1$ because x_3 is the single worst alternative at stage t_2. Hence, $C\left(\{x_1, x_2\}, S^{SWE}\right) = x_1$ and $C\left(\{x_2, x_3\}, S^{SWE}\right) = x_2$ necessarily imply $C\left(\{x_1, x_3\}, S^{SWE}\right) = x_1$ so that the choice function $C\left(\mathbf{X}, S^{SWE}\right)$ is always transitive. In order to show that may violate WARP, we provide the following example.

Example 5 *Let the set of alterntives be* $\mathbf{X} = \{x_1, x_2, x_3\}$ *and let the alternatives have the following aspects:* $x_1 = (1, 0, 0, ..)$, $x_2 = (0, 1, 1, ..)$ *and* $x_3 = (0, 1, 0, ..)$. *In accordance with SWE,* $C\left(\{x_1, x_2\}, S^{SWE}\right) = x_1$ *and* $C\left(\{x_1, x_1, x_3\}, S^{SWE}\right) = x_2$, *i.e., the WARP is violated.*

■

Proposition 5.3 states the optimal choice of a selection procedure guarantees only the transitivity of the resulting choices (of alternatives), but not the WARP. This seemingly irrational choice function arises because the decision-maker chooses an alternative indirectly: first, he chooses a procedure, and this choice satisfies WARP; then, the chosen procedure selects an alternative for the decision-maker.

Contrary to SWE, selection procedures SAT and EBA lead to random choice functions because $g^*(t_0, G, B) < G$ (in case of SAT) or $b^*(t_0, G, B) < B$ (in case of EBA) at a certain stage t_0. It can be shown that both SAT and EBA lead to random choice functions that are monotone (see Tversky, 1972), i.e., the probability that a given alternative is chosen from a choice set does not decrease if some other alternatives are removed from the set. This monotonicity property is often used for stochastic choice functions as a rationality criterion instead of WARP. Hence, in the limiting cases when $q_H = 1$ or $q_L = 0$, the choice function of a maximizing decision-maker looks rational. However, if $0 < q_L < q_H < 1$, i.e., when aspects do not fully reveal the types, his choice function violates WARP.

6 Conclusion

In this paper, I build a procedural model of choice based on CBE-model. The main ingredient of the model is the inability of a decision-maker to evaluate his objective function. This assumption makes it impossible to evaluate values of each alternative from the choice set and choose one with the highest value. Instead, the decision-maker uses aspects, which provide imperfect information about the value of the objective function that each alternative yields.

The decision-maker maximizes the probability of choosing the best alternative if he follows a specific procedure, which we call 'Single Worst Elimination'. In accordance with SWE, only the single worst alternative must be eliminated in each selection stage. This procedure is optimal independently of the size of a choice set, of the number of high type alternatives in it, and the accuracy of aspects (provided this accuracy is not perfect). The choice function induced by SWE is always transitive, but may violate the Weak Axiom of Revealed Preferences. When aspects become fully informative, then either 'Saticficing' or 'Elimination-By-Aspects' becomes optimal.

The main assumption that drives these results is the inability to evaluate an objective function for all alternatives. As a result, a decision-maker uses a procedure to select an alternative. In order to select a procedure, he constructs another, secondary objective function which is defined over the set of available procedures. Since the choice of a procedure is also a choice problem, the inability to evaluate an objective function must be applied to the secondary objective function as well. I have shown that the optimal selection procedure can be chosen without such evaluations. Hence, no procedure is needed in order to choose an optimal procedure.

I also make some other assumptions. First, every aspect corresponds to a single selection stage. This assumption is not essential and can be relaxed with redefining aspects as follows. A set of all aspects of an alternative that can be used in a certain selection stage defines a single generalized aspect of the alternative. The inability to evaluate objective functions requires that neither of these generalized aspects include all aspects. As a result of this modification, a decision-maker observes one generalized aspect in each stage.

Second, the decision-maker minimizes the type-II error. This assumption is equivalent to the assumption that alternatives are binary. In general, each alternative from the choice set is characterized by its rank within the set. Several secondary objective functions can be used. A decision-maker can either minimize the expected rank of the chosen alternative, or minimize the probability of choosing the lowest rank, or, as in the current model, maximize the probability that the chosen alternative has the highest rank. It can be shown that maximization of the expected objective function is equivalent to maximization of the expectation of a certain strictly decreasing function of the chosen rank. This function depends on the values of all alternatives from the choice set. Therefore, in order to accommodate non-binary types, the optimal selection procedure must maximize the expectation of any strictly decreasing function of the rank. The first order stochastic dominance criterion used in the current model, together with the monotonicity of that function, seems to be a powerful tool in generalizing the model in this direction.

Third, the number of aspects is infinite. This assumption assures that SWE is feasible. For finite but a sufficiently large number of aspects, the optimal selection procedure looks like SWE when there are still many stages ahead. Only when the number of remaining stages is small, the decision-maker has incentives to eliminate more than one bad alternative. In the last stage, only one alternative has to be selected, preferably a good alternative. Hence, SWE can be considered as being asymptotically optimal in the sense that when the number of aspects increases, the optimal procedure uses SWE in more initial stages of the selection.

Fourth, all aspects are equally informative. Without this assumption, the optimal procedure will depend on aspects' accuracies, and a decision-maker will have to know them (or, their prior distribution at least). As in the main assumption, I argue that this knowledge is unavailable to him so that the assumption that aspects are equally informative is a neutral ignorance assumption.

The model can be used in explaining different attitudes towards risk. Let us consider an expected utility maximizer, who does not know the prior distribution of states of the world. If we assume that he uses his Bernoulli utility functions as aspects, then he has to sequentially eliminate alternatives that yield low instantaneous utilities in each state of the world. In other words, he puts larger weights on the lowest levels of his utility. Hence, the model has an intrinsic aversion towards risk, which does not require concavity of the Bernoulli utility function.

The model can also be useful in analyzing humans' reference dependence. In accordance with Prospect Theory (Kahneman and Tversky, 1979), preferences of an individual are subject to a reference point that he uses. By using different alternatives from his choice set as reference points, an individual gets generally different subjective preferences. Each such preference relation can be viewed as an aspect. A choice function of such a decision-maker will be fully determined by the selection procedure that he follows.

7 Appendix to chapter 5

Proof of Proposition 5.1.

The proof is done by induction. We assume (induction assumption) that a selection procedure $S(g^*, b^*)$ is optimal for all $N \leq K$ so that the properties (a)-(d) from the proposition hold. It is easy to see that for $N_t = 1$, the only feasible selection procedure is $(g, b) = (G, B)$, which is trivially optimal and satisfies the properties (a)-(c) of Proposition 5.1. The property (d) of Proposition 5.1 is also satisfied as $U(S, 1, N_H) = N_H$ so that $U^*(1, N_H) \equiv N_H$ strictly increases in N_H. Thus, the induction assumption is satisfied for $K = 1$. Now, suppose that at a stage t of the selection:

a) the number of alternatives is $Nt = K + 1$;

b) the prior distribution of the number of high type alternatives N_t^H is $p_t(z)$, i.e., $p_t(z) \equiv \Pr(N_t^H = z)$;

c) the numbers of good and bad alternatives are G_t and $B_t = N_t - G_t$ correspondingly;

d) the decision-maker selects g_t good alternatives and b_t bad alternatives so that $g_t + b_t \leq N_t - 1 = K$.

Under these assumptions, the posterior distribution of the number of high type alternatives N_{t+1}^H at stage $(t+1)$ is:

$$p_{t+1}\left(N_{t+1}^H = z \mid N_t, N_t^H, G_t, g_t, b_t\right) \equiv \sum_{N_t^H=0}^{N} p_t\left(N_t^H\right) \Pr\left(N_{t+1}^H = z \mid N_t, N_t^H, G_t, g_t, b_t\right)$$

where $\Pr\left(N_{t+1}^H = z \mid N_t, N_t^H, G_t, g_t, b_t\right)$ is the probability that exactly z high type alternatives will be among the g_t good and b_t bad alternatives selected. As $g_t + b_t \leq K$, the number of alternatives at all stages after stage t does not exceed K and, by the induction assumption, $S(b^*, g^*)$ is optimal. Hence, after selecting g_t good alternatives and b_t bad alternatives in stage t, the decision-maker follows $S(b^*, g^*)$. The probability that a high type alternative will be finally selected depends on t, N_t, G_t, g_t, and b_t, and equals to:

$$\begin{aligned}
\hat{U}(t, N_t, G_t, g_t, b_t) &\equiv \sum_{z=0}^{z=g_t+b_t} p_{t+1}\left(N_{t+1}^H = z \mid N_t, N_t^H, G_t, g_t, b_t\right) U(S^*, g_t + b_t, z) \\
&= 1 - \sum_{z=0}^{z=g_t+b_t-1} F_{t+1}\left(w \mid N_t, N_t^H, G_t, g_t, b_t\right) \times \\
&\quad \times \left(U(S^*, g_t + b_t, z+1) - U(S^*, g_t + b_t, z)\right)
\end{aligned}$$

where $F_{t+1}\left(w \mid N_t, N_t^H, G_t, g_t, b_t\right) \equiv \sum_{z \leq w} p_{t+1}\left(N_{t+1}^H = z \mid N_t, N_t^H, G_t, g_t, b_t\right)$ is the corresponding cumulative distribution function. By the induction assumption, $U(..., z+1) - U(..., z) > 0$. Therefore, if the distribution

$F_{t+1}\left(w \mid N_t, N_t^H, G_t, g_t', b_t'\right)$ first-order stochastically dominates the distribution $F_{t+1}\left(w \mid N_t, N_t^H, G_t, g_t, b_t\right)$ then $\hat{U}\left(t, N_t, G_t, g_t', b_t'\right) > \hat{U}\left(t, N_t, G_t, g_t, b_t\right)$.

In the rest of the proof, we use this first-order stochastic dominance criterion in order to show that $\hat{U}\left(t, N_t, G_t, g^*, b^*\right) > \hat{U}\left(t, N_t, G_t, g', b'\right)$for all $(g, b) \neq \left(g^*\left(t, G_t, B_t\right), b^*\left(t, G_t, B_t\right)\right)$. We split this proof into five steps. In step one, we derive explicit expression for $F_{t+1}\left(w \mid N_t, N_t^H, G_t, g_t, b_t\right)$. In steps two, three and four, we derive properties (a)-(c) of Proposition 5.1. for any G_t and arbitrary prior distribution $p_t\left(z\right)$. In particular, in step two we show that $g^* = G_t$ for $b_t = 0$, whereas in step three we show that $g^* = G_t$ also for $b_t > 0$. Finally, in step four we show that first, it is either $b^* = 0$ or $b^* = B_t$, and second, that $b^* = B_t$ for $B_t \geq 2$. In step five, we prove part (d) of Proposition 5.1., which ends the induction arguments. The following notations are used throughout the proof:

$$\begin{aligned}
\alpha &\equiv \frac{q_H\left(1-q_L\right)}{q_L\left(1-q_H\right)} = 1 + \frac{q_H - q_L}{q_L\left(1-q_H\right)} \in (1, \infty) \\
Q_t &\equiv \frac{C_{N_t}^{G_t}}{C_{N_t}^{N_t^H} \sum_y C_{N_t^H}^{y} C_{N_t - N_t^H}^{G_t - y} \alpha^y}
\end{aligned}$$

where, here and after, C_y^x is the binomial coefficient $\binom{y}{x}$, all summation indices implicitly take all integer values, and the binomial coefficients are assumed to be zero if they are not defined for given values of its entries.

Step one. In order to derive the distribution $p_{t+1}\left(N_{t+1}^H = z \mid N_t, N_t^H, G_t, g_t, b_t\right)$, let assume that in stage t exactly y out of G_t good alternatives are of high type, and the remaining $G_t - y$ good alternatives are of low type. Selecting a sample (g_t, b_t) yields the following probability of having exactly z high type alternatives among the $g_t + b_t$ alternatives selected

$$\Pr\left(N_{t+1}^H = z \mid N_t, N_t^H, G_t, g_t, b_t, y_t\right) = \frac{\sum\limits_n C_y^n C_{N_t^H - y}^{z-n} C_{G_t - y}^{g_t - n} C_{N_t - N_t^H - G_t + y}^{b_t - z + n}}{C_{G_t}^{g_t} C_{N_t - G_t}^{b_t}}$$

Now, y follows the distribution $\Pr\left(y_t = z \mid N_t, N_t^H, G_t\right)$, which is:

$$\begin{aligned}
&= \frac{C_{N_t^H}^{y}\left(q_H\right)^y\left(1-q_H\right)^{N_t^H - y} C_{N_t - N_t^H}^{G_t - y}\left(q_L\right)^{G_t - y}\left(1-q_L\right)^{N_t - N_t^H - G_t + y}}{\sum\limits_y C_{N_t^H}^{y}\left(q_H\right)^y\left(1-q_H\right)^{N_t^H - y} C_{N_t - N_t^H}^{G_t - y}\left(q_L\right)^{G_t - y}\left(1-q_L\right)^{N_t - N_t^H - G_t + y}} \\
&= \frac{C_{N_t^H}^{y} C_{N_t - N_t^H}^{G_t - y} \alpha^y}{\sum\limits_y C_{N_t^H}^{y} C_{N_t - N_t^H}^{G_t - y} \alpha^y}
\end{aligned}$$

Then, taking expectations of the probability $\Pr\left(N_{t+1}^H = z \mid N_t, N_t^H, G_t, g_t, b_t, y_t\right)$ with respect to y, yields the following expression for $p_{t+1}\left(N_{t+1}^H = z \mid N_t, N_t^H, G_t, g_t, b_t\right)$:

$$\begin{aligned}
& \sum_y \Pr\left(N_{t+1}^H = z \mid N_t, N_t^H, G_t, g_t, b_t, y_t\right) \times \\
& \times \Pr\left(y_t = y \mid N_t, N_t^H, G_t\right) \\
= & \frac{\sum_y \sum_n C_{N_t^H}^{y} C_{N_t - N_t^H}^{G_t - y} C_y^n C_{N_t^H - y}^{z-n} C_{G_t - y}^{g_t - n} C_{N_t - N_t^H - G_t + y}^{b_t - z + n} \alpha^y}{C_{G_t}^{g_t} C_{N_t - G_t}^{b_t} \sum_y C_{N_t^H}^{y} C_{N_t - N_t^H}^{G_t - y} \alpha^y} \\
= & \; Q_t \sum_n C_{g_t}^n C_{b_t}^{z-n} \sum_y C_{G_t - g_t}^{y-n} C_{N_t - G_t - b_t}^{N_t^H - y - z + n} \alpha^y
\end{aligned}$$

and for $F_{t+1}\left(w \mid N_t, N_t^H, G_t, g_t, b_t\right)$:

$$F_{t+1}\left(w \mid N_t, N_t^H, G_t, g_t, b_t\right) = Q_t \sum_n \sum_y \sum_{z \le w} C_{h_t}^n C_{b_t}^{z-n} C_{G_t - g_t}^{y-n} C_{N_t - G_t - b_t}^{N_t^H - y - z + n} \alpha^y$$

Step two. (derivation of $g^* = G_t$ for $b_t = 0$). Suppose that the decision-maker does not select bad alternatives in stage t, i.e., $b_t = 0$, and selects $g_t \le G_t - 1$, i.e., not all, good alternatives. The corresponding type's distribution in stage $t+1$ is, by definition, $F_{t+1}\left(w \mid N_t, N_t^H, G_t, g_t, 0\right)$. However, he could get a better distribution (here and after, in terms of the first-order stochastic dominance) if he uses the following modified selection procedure $\tilde{S}$.

Let the decision-maker select $(g_t + 1)$good alternatives, i.e., one good alternative more. This will induce the distribution $p_{t+1}\left(N_{t+1}^H = z \mid N_t, N_t^H, G_t, g_t + 1, 0\right)$. Then, let him wait until at a certain stage, $t + m$, there are exactly g_t good alternatives and exactly one bad alternative. In that stage, let the decision-maker select only g_t good alternatives. This modified selection procedure $\tilde{S}$ yields $N_{t+1}^H = N_{t+m}^H$ by construction and induces the distribution $p_{t+m+1}\left(N_{t+m+1}^H = z \mid h_t + 1, N_{t+1}^H, g_t, g_t, 0\right)$. Taking expectations of the latter probability with respect to N_{t+1}^H, which is distributed in accordance with $p_{t+1}\left(N_{t+1}^H = z \mid N_t, N_t^H, G_t, g_t + 1, 0\right)$, yields the following distribution $\tilde{F}_{t+m+1}(w)$of the number of high type alternatives N_{t+m+1}^H among h_t alternatives in stage $t + m + 1$:

$$\begin{aligned}
\tilde{F}_{t+m+1}(w) \equiv & \sum_{z \le w} \sum_\mu p_{t+m+1}\left(N_{t+m+1}^H = z \mid g_t + 1, \mu, g_t, g_t, 0\right) \times \\
& \times p_{t+1}\left(N_{t+1}^H = \mu \mid N_t, N_t^H, G_t, g_t + 1, 0\right)
\end{aligned}$$

$$\tilde{F}_{t+m+1}(w) = Q_t \sum_{z\le w}\sum_{\mu=z}^{z+1} \frac{C_\mu^z C_{g_t+1-\mu}^{g_t-z}\alpha^{z+1-\mu}}{\mu+(g_t+1-\mu)\alpha} \sum_y C_{g_t+1}^{\mu} C_{G_t-g_t-1}^{y-\mu} C_{N_t-G_t}^{N_t^H-y}\alpha^y$$
$$= Q_t \sum_y C_{N_t-G_t}^{N_t^H-y}\alpha^y \left(\begin{array}{c} \sum_{z\le w} \frac{(g_t+1-z)C_{g_t+1}^{z} C_{G_t-g_t-1}^{y-z}\alpha}{z+(g_t+1-z)\alpha} + \\ + \sum_{z\le w} \frac{(z+1)C_{g_t+1}^{z+1} C_{G_t-g_t-1}^{y-z-1}}{z+1+(g_t-z)\alpha} \end{array} \right)$$
$$= Q_t \sum_y C_{N_t-G_t}^{N_t^H-y}\alpha^y \left(\begin{array}{c} \sum_{z\le w} C_{g_t+1}^{z} C_{G_t-g_t-1}^{y-z} + \\ + \frac{(w+1)C_{g_t+1}^{w+1} C_{G_t-g_t-1}^{y-w-1}}{w+1+(g_t-w)\alpha} \end{array} \right)$$

It can be seen now that the distribution $\tilde{F}_{t+m+1}(w)$ dominates $F_{t+1}\left(w \mid N_t, N_t^H, G_t, g_t, 0\right)$ as

$$D_0(w) \equiv \frac{1}{Q_t}\left(F_{t+1}\left(w \mid N_t, N_t^H, G_t, g_t, 0\right) - \tilde{F}_{t+m+1}(w)\right)$$
$$= \sum_y C_{N_t-G_t}^{N_t^H-y}\alpha^y \left(\begin{array}{c} \sum_{z\le w} C_{g_t}^{z} C_{G_t-g_t}^{y-z} - \sum_{z\le w} C_{g_t+1}^{z} C_{G_t-g_t-1}^{y-z} - \\ - \frac{(w+1)C_{g_t+1}^{w+1} C_{G_t-g_t-1}^{y-w-1}}{w+1+(g_t-w)\alpha} \end{array} \right)$$
$$= \sum_y C_{N_t-G_t}^{N_t^H-y}\alpha^y \left(C_{g_t}^{w} C_{G_t-g_t-1}^{y-w-1} + \frac{(w+1)C_{g_t+1}^{w+1} C_{G_t-g_t-1}^{y-w-1}}{w+1+(g_t-w)\alpha} \right)$$
$$= \frac{(g_t-w)C_{g_t}^{w}(\alpha-1)}{w+1+(g_t-w)\alpha} \sum_y C_{G_t-g_t-1}^{y-w-1} C_{N_t-G_t}^{N_t^H-y}\alpha^y > 0$$

for all $w \le g_t - 1$. Thus, if $b_t = 0$, an optimal selection procedure requires $g(t, G, B) = G$.

Step three. (derivation of$g^* = G_t$ for $b_t > 0$). Suppose next that the decision-maker does select bad alternatives in stage t, i.e., $b_t > 0$, and also selects $g_t \le G_t - 1$, i.e., not all, good alternatives. This induces the distribution $F_{t+1}\left(w \mid N_t, N_t^H, G_t, g_t, b_t\right)$ in stage $t+1$. However, he could get a better distribution if he replaces one bad alternative with one good alternative. Defining

$$D_b(w) \equiv \frac{1}{Q_t}\left(F_{t+1}\left(w \mid N_t, N_t^H, G_t, g_t, b_t\right) - F_{t+1}\left(w \mid N_t, N_t^H, G_t, g_t+1, b_t-1\right)\right)$$

yields:

$$\begin{aligned}
D_b(w) &= \sum_n \sum_y \alpha^y \sum_{z\le w} \begin{pmatrix} C^n_{g_t} C^{z-n}_{b_t} C^{y-n}_{G_t-g_t} C^{N^H_t-y-z+n}_{N_t-G_t-b_t} - \\ -C^n_{g_t+1} C^{z-n}_{b_t-1} C^{y-n}_{G_t-g_t-1} C^{N^H_t-y-z+n}_{N_t-G_t-b_t+1} \end{pmatrix} \\
&= \sum_n \sum_y \alpha^{y+1} \sum_{z\le w} \begin{pmatrix} C^n_{g_t} C^{z-n}_{b_t-1} C^{y-n}_{G_t-g_t-1} C^{N^H_t-y-z+n-1}_{N_t-G_t-b_t} - \\ -C^n_{g_t} C^{z-n-1}_{b_t-1} C^{y-n}_{G_t-g_t-1} C^{N^H_t-y-z+n}_{N_t-G_t-b_t} \end{pmatrix} - \\
&\quad -\sum_n \sum_y \alpha^y \sum_{z\le w} \begin{pmatrix} C^n_{g_t} C^{z-n}_{b_t-1} C^{y-n}_{G_t-g_t-1} C^{N^H_t-y-z+n-1}_{N_t-G_t-b_t} - \\ -C^n_{g_t} C^{z-n-1}_{b_t-1} C^{y-n}_{G_t-g_t-1} C^{N^H_t-y-z+n}_{N_t-G_t-b_t} \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
&= (\alpha-1) \sum_n C^n_{g_t} \sum_y C^{y-n}_{G_t-g_t-1} \alpha^y \times \\
&\quad \times \left(\sum_{z\le w} C^{z-n}_{b_t-1} C^{N^H_t-y-z+n-1}_{N_t-G_t-b_t} - \sum_{z\le w-1} C^{z-n}_{b_t-1} C^{N^H_t-y-z+n-1}_{N_t-G_t-b_t} \right) \\
&= (\alpha-1) \sum_n C^n_{g_t} C^{w-n}_{b_t-1} \sum_y C^{y-n}_{G_t-g_t-1} C^{N^H_t-y-w+n-1}_{N_t-G_t-b_t} \alpha^y > 0.
\end{aligned}$$

Thus, combining steps two and three, we have shown that an optimal selection procedure is such that $g^*(t, G, B) = G$.

Step four. (derivation of b^*). Suppose that the decision-maker, in addition to all good alternatives, also selects $b_t \le N_t - G_t - 1$, i.e., not all, bad alternatives. This induces the distribution $F_{t+1}\left(w \mid N_t, N^H_t, G_t, G_t, b_t\right)$ in stage $t+1$. Let us consider the following deviation $\hat{S}$ from this selection procedure. Let the decision-maker select l_t+1 bad alternatives, i.e., one bad alternative more. This will induce the distribution $p_{t+1}\left(N^H_{t+1} = z \mid N_t, N^H_t, G_t, G_t, b_t+1\right)$. Then, let wait until, at a certain stage $t+m$, there are exactly $G_t + b_t$ good alternatives and exactly one bad alternative. In that stage, let the decision-maker select only $G_t + b_t$ good alternatives. This modified selection procedure $\hat{S}$ yields$N^H_{t+1} = N^H_{t+m}$ by construction and induces the distribution $p_{t+m+1}\left(N^H_{t+m} = z \mid G_t + b_t + 1, N^H_{t+1}, G_t + b_t, G_t + b_t, 0\right)$at stage $t+m+1$. Taking expectations of the latter probability with respect to N^H_{t+1}distributed in accordance with $p_{t+1}\left(N^H_{t+1} = z \mid N_t, N^H_t, G_t, G_t, b_t+1\right)$ yields the following distribution $\hat{F}_{t+m+1}(w)$ of the number of high type alternatives N^H_{t+m+1} among $G_t + b_t$ alternatives in stage $t+m+1$:

$$\begin{aligned}
\hat{F}_{t+m+1}(w) \equiv& \sum_{z\le w} \sum_\mu p_{t+m+1}\left(N^H_{t+m+1} = z \mid G_t + b_t + 1, \mu, G_t + b_t, G_t + b_t, 0\right) \times \\
&\times p_{t+1}\left(N^H_{t+1} = \mu \mid N_t, N^H_t, G_t, G_t, b_t + 1\right)
\end{aligned}$$

$$\begin{aligned}
\hat{F}_{t+m+1}(w) &= Q_t \sum_{z\le w}\sum_{\mu=z}^{z+1} \frac{C_{\mu}^{z} C_{G_t+b_t+1-\mu}^{G_t+b_t-z}\alpha^{z+1-\mu}}{\mu+(G_t+b_t+1-\mu)\alpha} \sum_{y} C_{G_t}^{y} C_{b_t+1}^{\mu-y} C_{N_t-G_t-b_t-1}^{N_t^H-\mu}\alpha^{y} \\
&= Q_t \sum_{y} C_{G_t}^{y}\alpha^{y} \sum_{z\le w} \frac{(G_t+b_t+1-z)\, C_{g_t+1}^{z} C_{G_t-g_t-1}^{y-z}\alpha}{z+(G_t+b_t+1-z)\alpha} + \\
&\quad + Q_t \sum_{y} C_{G_t}^{y}\alpha^{y} \sum_{z\le w} \frac{(z+1)\, C_{b_t+1}^{z+1-y} C_{N_t-G_t-b_t-1}^{N_t^H-z-1}}{z+1+(G_t+b_t-z)\alpha} \\
&= Q_t \sum_{y} C_{G_t}^{y}\alpha^{y} \left(\begin{array}{c} \sum_{z\le w} C_{b_t+1}^{z-y} C_{N_t-G_t-b_t-1}^{N_t^H-z} + \\ + \frac{(w+1)C_{b_t+1}^{w+1-y} C_{N_t-G_t-b_t-1}^{N_t^H-w-1}}{w+1+(G_t+g_t-w)\alpha} \end{array} \right)
\end{aligned}$$

In order to compare distributions and $F_{t+1}\left(w \mid N_t, N_t^H, G_t, G_t, b_t\right)$ and $\hat{F}_{t+m+1}(w)$, we define

$$D_g(w, G_t, b_t) \equiv \frac{w+1+(G_t+b_t-w)\alpha}{Q_t C_{N_t-G_t-b_t-1}^{N_t^H-w-1}} \left(F_{t+1}\left(w \mid N_t, N_t^H, G_t, G_t, b_t\right) - \hat{F}_{t+m+1}(w)\right)$$

Then

$$\begin{aligned}
D_g(w, G_t, b_t) &= \sum_{y} C_{G_t}^{y}\alpha^{y} \left(\begin{array}{c} \frac{w+1+(G_t+b_t-w)\alpha}{C_{N_t-G_t-b_t-1}^{N_t^H-w-1}} \times \\ \times \sum_{z\le w}\left(C_{b_t}^{z-y} C_{N_t-G_t-b_t}^{N_t^H-z} - C_{b_t+1}^{z-y} C_{N_t-G_t-b_t-1}^{N_t^H-z} \right) \end{array} \right) \\
&\quad - \sum_{y} C_{G_t}^{y}\alpha^{y}(w+1)\, C_{b_t+1}^{w+1-y} \\
&= \sum_{y} C_{G_t}^{y}\alpha^{y} \left((w+1+(G_t+b_t-w)\alpha)\, C_{b_t}^{w-y} - (w+1)\, C_{b_t+1}^{w+1-y} \right) \\
&= \sum_{y} (G_t+b_t-w)\, C_{G_t}^{y} C_{b_t}^{w-y}\alpha^{y+1} \sum_{y} (w+1)\, C_{G_t}^{y} C_{b_t}^{w+1-y}\alpha^{y} \\
&= \sum_{y} \left((G_t+b_t-w)\, C_{G_t}^{y-1} - (w+1)\, C_{G_t}^{y} \right) C_{b_t}^{w-y+1}\alpha^{y}
\end{aligned}$$

It is easily seen that $D_g(w, 0, b_t) = (b_t - w)\, C_{b_t}^{w}\alpha - (w+1)\, C_{b_t}^{w+1} = b_t C_{b_t-1}^{w}(\alpha - 1) > 0$ for all $b_t \ge w+1 \ge 1$, and that $D_g(w, 0, 0) = 0$. In addition, $D_g(w, G_t, b_t)$ can be written recursively as

$$D_g(w, G_t+1, b_t) = D_g(w, G_t, b_t) + \alpha D_g(w-1, G_t, b_t)$$

Therefore, $D_g(w, G_t, 0) = 0$ and $D_g(w, G_t, b_t) > 0$ for all $b_t \ge 1$ by induction.

This implies that in stage t, any selection procedure (G_t, b_t) with $b_t = 1, .., N_t - G_t - 1$ is strictly dominated by the selection procedure $\hat{S}$. Hence, if some bad alternatives are eliminated, then all the other bad alternatives must also be eliminated, or, stated differently, if some bad alternatives are selected, then all the other bad alternatives must also be selected. In other words, all bad alternatives must be treated equivalently, i.e., either $b^*(t, G, B) = 0$ or $b^*(t, G, B) = B$.

Suppose now, that $G_t \leq N_t - 2$, i.e., at least two bad alternatives are available. Then, selecting $(G_t, 0)$ is not better than selecting $(G_t, 1)$, whereas selecting $(G_t, 1)$ is strictly dominated by selecting $(G_t, 2)$. Hence, selecting only good alternatives is also strictly dominated by passing all alternatives provided $G_t \leq N_t - 2$, i.e., if $b^*(t, G, B) = B$ if $B \geq 2$.

Step five (monotonicity of $u\left(S^*, N, N^H\right)$). From steps 2, 3 and 4 it follows that it is only optimal to eliminate an alternative if it is a unique bad alternative at a stage. Suppose that the decision-maker eliminates such a unique bad alternative,that is, $(g_t, b_t) = (N_t - 1, 0)$ when $(G_t, B_t) = (N_t - 1, 1)$. Evaluating $F_{t+1}\left(w \mid N_t, N_t^H, N_t - 1, N_t - 1, 0\right)$ yields:

$$\begin{aligned} F_{t+1}\left(w \mid N_t, N_t^H, N_t - 1, N_t - 1, 0\right) &= \frac{N_t \sum_{z \leq w} C_{N_t-1}^{z} C_1^{N_t - z} \alpha^{z+1-N_t^H}}{C_{N_t}^{N_t^H}\left(N_t^H + \left(N_t - N_t^H\right)\alpha\right)} \\ &= \begin{cases} 0 & \text{if } w < N_t^H - 1 \\ \frac{N_t^H}{N_t^H + \left(N_t - N_t^H\right)\alpha} & \text{if } w = N_t^H - 1 \\ 1 & \text{if } w > N_t^H - 1 \end{cases} \end{aligned}$$

It is easy to see that so that $F_{t+1}\left(w \mid N_t, N_t^H + 1, N_t - 1, N_t - 1, 0\right) \leq F_{t+1}\left(w \mid N_t, N_t^H, N_t - 1, N_t - 1, 0\right)$. $F_{t+1}\left(w \mid N_t, N_t^H + 1, N_t - 1, N_t - 1, 0\right)$ stochastically dominates the distribution $F_{t+1}\left(w \mid N_t, N_t^H, N_t - 1, N_t - 1, 0\right)$ and, therefore, $u\left(S^*, N, N^H + 1\right) > u\left(S^*, N, N^H\right)$ for all $N \leq N_t$. This ends the induction arguments and the proof of Proposition 1. ■

6. Conclusion

Since the pioneering work of Akerlof (1970) 'information' has become an essential term in the economics vocabulary, and informational issues have been playing a crucial role in the research agenda. The main argument of most of the literature that followed Akerlof's paper is that many economic inefficiencies (i.e., adverse selection and moral hazard) are due to a lack of information. However, in recent years, it has become clear that the quantity and the quality of the information at disposal of an economic agent is not the only element to determine the quality of economic outcomes. Prior to making decisions, economic agents must interpret the information they possess and, therefore, the information processing phase is itself a key aspect of decision making.

In the current thesis I have considered three specific information processing limitations: limited attention, lack of communication and imperfect recall. The first typically refers to the fact that, quantitatively, more information requires more resources (i.e., time and money) to understand and correctly interpret information. The second limitation is due to the fact that, qualitatively, complex information is not easily transmittable between different agents. Finally, imperfect recall refers to a situation in which an individual has to carry out several successive actions but faces memory limitations. Accounting for such limitations has allowed me to shed light on the effect that imperfect information processing has on the functioning of markets (chapter 2) collective decision making (chapters 3 and 4) and individual decision making (chapter 5).

In chapter 2 I have investigated the equilibrium property of a monopsony market in which many senders compete for a vacancy posted by one receiver who has limited attention. I have found that, even when information is a disposable good (e.g., information in excess can be disregarded) there exists a negative relation between the amount of information (i.e., the total amount of applications sent) and the average value of information processed. This result is in line with theoretical and empirical findings suggesting the existence of information overload: large amount information may lead to undesired outcomes.

In chapter 3 I have analyzed the collective decision making of a hierarchical organization whose task is to select the best alternative from a feasible set. Information processing refers here to the fact that, for each agent, the evaluation of alternatives comes in the form of imprecise news about each alternative's quality. Interestingly enough, my findings suggest that the ways in which limited attention and lack of communication shape the optimal selection rule are very similar. Under both information processing limitations, potentially valuable information is generally disregarded in equilibrium. In case of limited attention it is generally optimal for the agent placed at lower hierachical level to disregard some news (both good and bad) as this prevents to degrade excessively the screening accuracy of the agent that operates at higher level. In case of lack of communication it is generally optimal for the first agent to disregard some bad news as mixing good and bad news in a single pool makes it impossible to distinguish between them later on and, therefore, decreases the probability of selecting the best alternative.

In chapter 4 I use the same model used in chapter 3 but I limit my attention only to lack of communication, and my analysis focuses here on the optimal ordering of heterogenous agents within the organization. I have found that, by letting the more accurate agent screen first, the organization overcomes the imperfections due to lack of communication and is as efficient as a hierarchy in which information flows are perfect. It is important to stress that I didn't account for the direct costs usually involved in any information-processing phase. As it is natural to assume that the cost of processing information increases with the accuracy with which information is processed, it might be the case that having higher accuracy at higher stages is optimal for the simple reason that, at later stages, less information is processed. Whenever such ordering is in place, the contribution of my analysis is that of identifying the indirect cost that might emerge due to a lack of internal communication.

In chapter 5 I analyze the property of the choice function of a decision-maker who is unable to evaluate his objective function. I took a procedural approach to decision-making by modeling choice as a selection procedure of Choice by Elimination (CBE) (Tversky (1972a,1972b)). This approach assumes that each alternative is viewed as a sequence of aspects that the alternative may or may not posses, and the decision-maker's inability to evaluate the objective function is then captured by the fact that the different aspects of each alternative cannot be aggregated into a single performance measure. Since in my model each aspect has the interpretation of being a good news about the quality of the corresponding alternative, the choice problem can be interpreted as one with imperfect recall: the decision-maker sequentially observes informative signals about the quality of each alternatives, but he does not remember the signals observed in previous stages. My findings that the optimal selection procedure (namely, the 'Single Worst Elimination' (SWE) procedure) induces a choice function that violates the Weak Axiom of Revealed Preferences has one important implication. Shall we conclude, from the violation of the Weak Axiom, that the decision-maker is not rational? The answer is no. Due to imperfect recall, the original decision problem of choosing between a set of alternatives looses its meaning as it is necessarily rephrased as an auxiliary decision problem where the initial set of alternatives is replaced by the set of available procedures that could be implemented. Since a rational decision-maker always implements the optimal procedure, the choice function whose domain is the set of available procedures never violates the Weak Axiom. In other words, as imperfect recall is assumed, the rationality of the D-M conveys in choosing the procedure which is most appropriate for his goal. Therefore the main massage of chapter 5 is that seemingly irrational choices may still be consistent with the concept of rationality.

7. Samenvatting (summary in Dutch)

Sinds het baanbrekende werk van Akerlof (1970) is de term 'informatie' uitgegroeid tot een essentieel begrip in het economische vocabulaire en is er een hele reeks vervolgonderzoek naar het onderwerp gedaan. Het voornaamste argument dat men tegenwoordig in de literatuur betreffende Akerlof's paper tegenkomt is, dat economische inefficiëntie (b.v. adverse selection en moral hazard) te wijten zijn aan het gebrek aan informatie. Echter heeft men recentelijk vastgesteld, dat de kwantiteit en kwaliteit van de informatie, die een individu ter beschikking heeft, niet de enige indicatoren zijn om de kwaliteit van een economisch product vast te kunnen stellen. Om een besluit te kunnen nemen, vertolkt ieder individu de verzamelde informatie, waarom gesteld kan worden, dat het proces van informatieverwerking zelf een sleutel element voor het nemen van beslissingen is.

In deze scriptie heb ik tevens drie specifieke beperkingen van informatie processen aan bod gebracht: beperkte attentie, tekort aan communicatie en de beperkingen in het herinneringproces. De eerste beperking verwijst naar het feit dat, kwantitatief meer informatie meer middelen, zoals tijd en geld, opeist om de informatie te kunnen begrijpen en interpreteren. De tweede beperking bespreekt de complicaties die optreden zodra kwalitatief complexe informatie overgedragen wordt tussen twee of meerdere individuen. Het slotargument beschrijft een situatie, waarin een individu enkele opeenvolgende acties moet ondernemen waarbij hij op een beperking van het herinneringsvermogen stoot. Door rekening te houden met de genoemde beperkingen, was het mogelijk het effect te belichten, dat imperfecte verwerking van informatie op het functioneren van de markt (hoofdstuk2), het collectieve nemen van beslissingen (hoofdstuk 3 en 4) en individuele nemen van een besluit (hoofdstuk5), heeft.

In hoofdstuk 2 heb ik het evenwichtige eigendom van een monopsony markt onderzocht, waarin tal van afzenders concurreren om een vacature geplaatst door een ontvanger met beperkte attentie. Het gevonden resultaat beschreef dat, ook al is informatie een beschikbaar goed, er een negatieve relatie tussen de hoeveelheid informatie (b.v. het totale aantal beschikbare sollicitaties) en de doorsnee waarde gehecht aan de verwerkte informatie bestaat. Dit resultaat houdt verband met de theoretische en empirische bevindingen betreffende 'information overload', dat een groot aantal informatie tot ongehoopte resultaten leidt.

In hoofdstuk 3 heb ik het proces van collectieve besluitneming van een hiërarchische organisatie geanalyseerd, wiens taak het is het beste alternatief uit een reeks te vinden. Informatie verwerking refereert in dit geval naar het feit dat, de evaluatie van alternatieven voor elke speler in de vorm van onprecies nieuws over de kwaliteit van elk alternatief verschijnt. Interessant was dat, mijn bevindingen suggereren dat de manier waarop beperkte attentie en een gebrek aan communicatie de richtlijn voor een optimale selectie vormen, soortgelijk zijn. In beide beperkingen van het verwerken van informatie werd potentieel waardevolle informatie over het algemeen in een equilibrium buiten beschouwing gelaten. In het geval van het gebrek aan attentie is het over het algemeen het beste voor een speler op lager hiërarchisch niveau om, enkele informatie (goede en slechte) buiten beschouwing te laten, om te voorkomen dat, de nauwkeurigheid in het verwerken van informatie van een hoger geplaatste speler niet verlaagd word. In het geval van een gebrek aan communicatie is het over het algemeen voor de eerste speler het beste om slechte informatie buiten beschouwing te laten. Het samenvoegen van goede en slechte informatie in een pool maakt het onmogelijk op een later tijdstip een onderscheid tussen beiden te maken, waardoor de waarschijnlijkheid het beste alternatief te vinden kleiner wordt.

In hoofdstuk 4 heb ik het zelfde model gebuikt als in hoofdstuk 3. Echter heb ik me hier beperkt tot een tekort aan communicatie en mijn analyse is gericht op een optimale ordening van heterogene spelers in een organisatie. Mijn conclusie was, dat als men betrouwbaardere speler eerst laat sorteren het probleem van een tekort aan communicatie binnen de organisatie overkomen kan worden, waardoor het net zo efficiënt werkt als binnen een hiërarchische organisatie, waarin de stroming van informatie als perfect beschreven kan worden. Belangrijk is tevens te betonen, dat ik geen rekening met de directe kosten van een informatie verwerking proces gehouden heb.

Hoofdstuk 5 is een analyse van de eigenschappen van een keuzefunctie van een besluitnemend individu, die niet in staat is, de objectieve functie te evalueren. In dit prospect heb ik voor een aanpak gekozen, die besluitneming in een model weergeeft waarin keuze door middel van het uitsluitprincipe bepaald wordt (CBE) (Tversky (1972a, 1972b)). Deze aanpak veronderstelds, dat elk alternatief gezien wordt als een aaneenschakeling van aspecten dat het alternatief wel of niet zou kunnen hebben. De besluitnemers onbekwaamheid de objectieve functie te evalueren is nu overgenomen door het feit dat, verschillende aspecten van elk alternatief niet samengevoegd kunnen worden tot een enkele indicator van functionering. Aangezien elk aspect in mijn model geïnterpreteerd kan worden als zijnde goed nieuws over de kwaliteit van een alternatief kan het probleem van de keuze geïnterpreteerd worden als een complicatie van imperfecte herinnering. Er is sprake van imperfecte herinnering wanneer een besluitnemer achtereenvolgens informatieve signalen over de kwaliteit van elk alternatief opvangt, maar echter eerdere signalen niet meer kan herinneren. Mijn conclusie dat de optimale selectie procedure (met name, de 'Single Worst Elimination' (SWE) procedure) een keuzefunctie omvat die de 'Weak Axiom of Revealed Preferences' veronachtzaamd, heeft een belangrijke kanttekening. Zou geconcludeerd kunnen worden, dat volgens de veronachtzaming van de Weak Axiom de besluitnemende speler niet rationeel is? Het antwoord is nee. Volgens de imperfecte herinnering verliest het originele probleem van besluitneming zijn eigenlijke betekenis. Omdat een rationele besluitnemer altijd de optimale procedure zal doorvoeren om het beste alternatief tussen een reeks te vinden, zal de keuzefunctie de Weak Axiom nooit veronachtzamen. De essentie van hoofdstuk 5 is tevens dat, schijnbaar irrationele keuzes ondanks het concept van rationaliteit kunnen blijven bestaan.

8. Bibliography

Abreu, D and A. Rubinstein, (1988). "The Structure of Nash Equilibrium in Repeated Games with Finite Automata", *Econometrica*, 56, 1259-1281.

Akerlof, G. (1970). "The Market of Lemons: Quality Uncertainty and the Market Mechanism". *Quarterly Journal of Economics*, 89, 488-500.

Anderlini L., (1991). "Communication, Computability, and Common Interest Games", Working Papers, St. John's College, Cambridge.

Anderlini L., and L. and Felli, (1992). "Incomplete Written Contracts", Working Papers, London School of Economics.

Bolton, P., and M. Dewatripont (1994). "The Firm as a Communication Network", *Quarterly Journal of Economics* 109, 809-839.

Bull C., and J. A. Ordover (1987). "Market Structure and Management Organizations", *RAND Journal of Economics* 18, 480-491.

Chen, H-C., J. W. Friedman, and J-F. Thisse (1997). "Boundedly Rational Nash Equilibrium: a Probabilistic Choice Approach", *Games and Economic Behavior* 18, 32-54.

Delfgaauw J., and R. A. J. Dur (2005). "Signaling and Screening of Workers' Motivation", Working Paper, Tinbergen Institute, Erasmus University Rotterdam (forthcoming in *Journal of Economic Behavior and Organization*).

Easley, D. and A. Rustichini, (1999). "Choice without Beliefs", *Econometrica* 67, 1157-1184.

Easley, D. and A. Rustichini, (2005). "Optimal Guessing: Choice in Complex Environments", *Journal of Economic Theory* 124, 1-21.

Gilboa, I. and D. Schmeidler, (1995). "Case-based Decision Theory", *Quarterly Journal of Economics* 110, 605-639.

Gilboa, I. and D. Schmeidler, (1997). "Cumulative Utility Consumer Theory", *International Economic Review* 38, 737-761.

Hendrikse, G.W.J., (1992). "Competition Between Architectures". *Annales d'Economie et de Statistique*, 25/26, 39-50.

Huber, J., J. Payne, and C. Puto, (1982). "Adding Asymmetrically Dominated Alternatives: Violation of Regularity and Similarity Hypothesis", *Journal of Consumer Research* 9, 90-98.

Ioannides, Y.M., (1987). "On the Architecture of Complex Organizations". *Economics Letters*, 25, 201-206.

Jacoby, J., Speller, D. E. and Kohn, C. A. (1974). 'Brand Choice Behavior as a Function of Information Load'. *Journal of Marketing Research* 11, 63-69.

Janssen, M. C. W. (2002) ."Catching Hipos: Screening, Wages, and Competing for a Job", *Oxford Economic Papers* 54 , 321-33.

Kahneman, D. and A. Tversky, (1979). Prospect theory: An Analysis of Decision under Risk, *Econometrica* 47, 263-292.

Kalai, G., A. Rubinstein and R. Spiegler, (2002). "Rationalizing Choice Functions by Multiple Rationales", *Econometrica* 70, 2481-2488.

Kalai E., and W. Stanford (1988). "Finite Rationality and Interpersonal Complexity in Repeated Games", *Econometrica*, 56, 397-410.

Keller, K. L. and Staelin, R. (1987). "Effects of Quality and Quantity of Information on Decision Effectiveness", *Journal of Consumer Research* 14, 200–213.

Koh, W. T. H., (1992a). "Variable Evaluation Costs and the Design of Fallible Hierarchies and Polyarchies". *Economics Letters*, 38, 313-318.

Koh, W. T. H., (1992b). "Human Fallibility and Sequential Decision Making. Hierarchy versus Polyarchy". *Journal of Economic Behavior and Organization* 18, 317-345.

Koh, W. T. H., (1993). "First-Mover Advantage and Organizational Structure". *Economics Letters*, 43,47-52.

Koh, W. T. H., (1994a). "Making Decisions in Committees. A Human Fallibility Approach". *Journal of Economic Behavior and Organization*, 23, 195-214.

Koh, W. T. H., (1994b). "Fallibility and Sequential Decision-Making". *Journal of Institutional and Theoretical Economics*, 150, 362-374.

Kraut, R. E., Sunder, S., Morris, J., Filer, D., and Cronin, M. (2002). "Markets for Attention:Will Postage for Email Help?", *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work,* 206–215.

Lambert R. A., (1986). "Executive Effort and Selection of Risky Projects". *The RAND Journal of Economics* 17, 77-88.

Lancaster, K., (1966). "A New Approach to Consumer Theory", *Journal of Political Economy* 74, 132-157.

Libowski, Z. (1975). "Sensory and Information Inputs Overload: Behavioral Effects", *Comprehensive Psychiatry* 16, 199–221.

Luce, R. D., and H. Raiffa (1957). "Games and Decisions". New York: John Wiley.

Malhotra, N. K. (1982). "Information Load and Consumer Decision Making", *Journal of Consumer Research* 8, 419–430.

Manzini P., and M. Mariotti (2005) "Rationalizing Boundedly Rational Choice: Sequential Rationalizability and Rational Shortlist Method", Mimeo, Queen Mary, University of London.

Marschak, J., and R. Radner (1972). "Economic Theory of Teams", New Haven, Conn.: Yale University Press.

Meagher, K,. and T. Van Zandt, (1998). "Managerial Costs for One-Shot Decentralized Information Processing", *Review of Economic Design* 3, 329-345.

Meyer, R. J., and E. J. Johnson (1989). "Information Overload and the Nonrobustness of Linear Models: a Comment on Keller and Staelin", *The Journal of Consumer Research* 15, 498-503.

Milgrom, P., (1981). "Good News and Bad News: Representation Theorems and Applications", *Bell Journal of Economics* 12, 380-391.

Miller, G. A. (1956). "The Magical number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information", *Psychological Review* 63, 81-97.

Moscarini G., and L. Smith (2001). "The Optimal Level of Experimentation", *Econometrica* 69, 1629-1644.

Moscarini G., and L. Smith (2002). "The Law of Large Demand for Information", *Econometrica* 70, 2351-2366.

Osborne, M. J., and A. Rubinstein, (1998). "Games with Procedurally Rational Players", *American Economic Review* 88, 834-847.

Radner, R., (1992), "Hierarchy: the Economics of Managing", *Journal of Economic Literature* 30, 1382-1415.

Radner, R., and T. Van Zandt, (1992). "Information Processing in Firms and Returns to Scale", *Annales d'economie et de statistique* 25/26, 265-298.

Radner, R., (1993), "The Organization of Decentralized Information Processing", *Econometrica* 61, 1109-1146.

Radner, R., and T. Van Zandt, (2001). "Real-Time Decentralized Information Processing and Returns to Scale", *Economic Theory* 17, 497-544.

Rosenthal, R. W., (1989). "A bounded-rationality Approach to the Study of Noncooperative Games", *International Journal of Game Theory* 18, 273-292.

Rubinstein, A., (1986). "Finite Automata Play the Repeated Prisoners' Dilemma", *Journal of Economic Theory*, 39, 83-96.

Rubinstein, A., (1988a). "Modeling Bounded Rationality", Cambridge MA: MIT press.

Rubinstein, A., (1988b). "Similarity and Decision-Making under Risk (Is There a Utility Theory Resolution to the Allais Paradox?)", *Journal of Economic Theory* 46, 145-153.

Sah, R. K., and J. E. Stiglitz (1985). "Human Fallibility and Economic Organization". *American Economic Review*, 75, 292-297.

Sah, R. K., and J. E. Stiglitz (1986). "The Architecture of Economic Systems: Hierarchies and Polyarchies". *American Economic Review*, 76, 716-727.

Sen A., (1993). "Internal Consistency of Choice", *Econometrica*, 61 495-521.

Simon, H., (1955). "A Behavioral Model of Rational Choice", *Quarterly Journal of Economics* 69, 99-118.

Spear, S., (1989). "Learning Rational Expectations under Computability Constraints". *Econometrica* 57, 889-910.

Tversky, A., (1969). "Intransitivities of Preferences". *Psychological Review* 76, 31-48.

Tversky, A., (1972a). "Choice by Elimination", *Journal of Mathematical Psychology* 9, 341-367.

Tversky, A., (1972b). "Elimination by Aspects: A theory of choice". *Psychological Review*, 79, 281-299.

Tversky, A., (1977). "Features of Similarity", *Psychological Review* 84, 327-352.

Tversky, A., and D. Kahneman (1981). "The Framing of Decisions and the Psychology of Choice", *Science* 211, 453-458.

Tversky, A., and D. Kahneman (1986). "Rational Choice and the Framing of Decisions", *Journal of Business* 59, 251-278.

Tversky, A., and E. Shafir (1992). "Choice under Conflict: the Dynamics of Deferred Decisions", *Psychological Science* 3, 358-361.

Van Zandt, T., (1997). "The Scheduling and Organization of Periodic Associative Computation: Essential Networks" *Review of Economic Design* 3,15-27.

Van Zandt, T., (1998). "The Scheduling and Organization of Periodic Associative Computation: Efficient Networks", *Review of Economic Design* 3, 93-127.

Van Zandt, T. (1999). "Real-Time Decentralized Information Processing as a Model of Organizations with Boundedly Rational Agents", *Review of Economic Studies* 66, 633-658.

Van Zandt, T. (2004). "Information Overload in a Network of Targeted Communication", *RAND Journal of Economics* 35, 542-560.

Visser, B., (2000). "Organizational Communication Structure and Performance", *Journal of Economic Behavior and Organization*, 42, 231-252.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

356. C.M. VAN DER HEIDE, *An economic analysis of nature policy.*
357. Y. HU, *Essays on labour economics: Empirical studies on wage differentials across categories of working hours, employment contracts, gender and cohorts.*
358. S. LONGHI, *Open regional labour markets and socio-economic developments: Studies on adjustment and spatial interaction.*
359. K.J. BENIERS, *The quality of political decision making: Information and motivation.*
360. R.J.A. LAEVEN, *Essays on risk measures and stochastic dependence: With applications to insurance and finance.*
361. N. VAN HOREN, *Economic effects of financial integration for developing countries.*
362. J.J.A. KAMPHORST, *Networks and learning*.
363. E. PORRAS MUSALEM, *Inventory theory in practice: Joint replenishments and spare parts control.*
364. M. ABREU, *Spatial determinants of economic growth and technology diffusion.*
365. S.M. BAJDECHI-RAITA, *The risk of investment in human capital*.
366. A.P.C. VAN DER PLOEG, *Stochastic volatility and the pricing of financial derivatives.*
367. R. VAN DER KRUK, *Hedonic valuation of Dutch Wetlands*.
368. P. WRASAI, *Agency problems in political decision making.*
369. B.K. BIERUT, *Essays on the making and implementation of monetary policy decisions*.
370. E. REUBEN, *Fairness in the lab: The effects of norm enforcement in economic decisions.*
371. G.J.M. LINDERS, *Intangible barriers to trade: The impact of institutions, culture, and distance on patterns of trade.*
372. A. HOPFENSITZ, *The role of affect in reciprocity and risk taking: Experimental studies of economic behavior.*
373. R.A. SPARROW, *Health, education and economic crisis: Protecting the poor in Indonesia.*
374. M.J. KOETSE, *Determinants of investment behaviour: Methods and applications of meta-analysis.*
375. G. MÜLLER, *On the role of personality traits and social skills in adult*

economic attainment.

376. E.H.B. FEIJEN, *The influence of powerful firms on financial markets.*
377. J.W. GROSSER, *Voting in the laboratory.*
378. M.R.E. BRONS, M*eta-analytical studies in transport economics: Methodology and applications*.
379. L.F. HOOGERHEIDE, *Essays on neural network sampling methods and instrumental variables.*
380. M. DE GRAAF-ZIJL, *Economic and social consequences of temporary employment.*
381. O.A.C. VAN HEMERT, *Dynamic investor decisions.*
382. Z. ŠAŠOVOVÁ, *Liking and disliking: The dynamic effects of social networks during a large-scale information system implementation.*
383. P. RODENBURG, *The construction of instruments for measuring unemployment.*
384. M.J. VAN DER LEIJ, *The economics of networks: Theory and empirics.*
385. R. VAN DER NOLL, *Essays on internet and information economics.*
386. V. PANCHENKO; *Nonparametric methods in economics and finance: dependence, causality and prediction.*
387. C.A.S.P. SÁ, *Higher education choice in The Netherlands: The economics of where to go.*
388. J. DELFGAAUW, *Wonderful and woeful work: Incentives, selection, turnover, and workers' motivation.*
389. G. DEBREZION, *Railway impacts on real estate prices.*
390. A.V. HARDIYANTO, *Time series studies on Indonesian rupiah/USD rate 1995 – 2005.*
391. M.I.S.H. MUNANDAR, *Essays on economic integration.*
392. K.G. BERDEN, *On technology, uncertainty and economic growth.*
393. G. VAN DE KUILEN, *The economic measurement of psychological risk attitudes.*
394. E.A. MOOI, *Inter-organizational cooperation, conflict, and change.*
395. A. LLENA NOZAL, *On the dynamics of health, work and socioeconomic status.*
396. P.D.E. DINDO, *Bounded rationality and heterogeneity in economic dynamic models*.
397. D.F. SCHRAGER, *Essays on asset liability modeling*.
398. R. HUANG, *Three essays on the effects of banking regulations*.
399. C.M. VAN MOURIK, G*lobalisation and the role of financial accounting information in Japan.*
400. S.M.S.N. MAXIMIANO, *Essays in organizational economics.*
401. W. JANSSENS, *Social capital and cooperation: An impact evaluation of a women's empowerment programme in rural India.*
402. J. VAN DER SLUIS, *Successful entrepreneurship and human capital.*
403. S. DOMINGUEZ MARTINEZ, *Decision making with asymmetric*

information.

404. H. SUNARTO, *Understanding the role of bank relationships, relationship marketing, and organizational learning in the performance of people's credit bank.*

405. M.Â. DOS REIS PORTELA, *Four essays on education, growth and labour economics.*