TI 2004-077/1

Tinbergen Institute Discussion Paper

Information Overload in Multi-Stage Selection Procedures

*Stefano Ficco**

Vladimir A. Karamychev

Faculty of Economics, Erasmus Universiteit Rotterdam.

** Tinbergen Institute.*

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Please send questions and/or remarks of non-scientific nature to driessen@tinbergen.nl.
Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

Information Overload in Multi-Stage Selection Procedures

Stefano Ficco

(Tinbergen Institute Rotterdam)

and

Vladimir A. Karamychev

(Erasmus University Rotterdam)

This version: 06-2004

Abstract: The paper studies information processing imperfections in a fully rational decision-making network. It is shown that imperfect information transmission and imperfect information acquisition in a multi-stage selection game yield information overload. The paper analyses the mechanisms responsible for a seeming bounded rational behavior of the network and shows their similarities and distinctions. Two special cases of filtering selection procedures are investigated, where the overload takes its most limiting forms. The model developed in the paper can be applied both to organizations and to individuals. It can serve as a rational foundation for bounded rationality.

We are very grateful to Maarten Janssen for many discussions and helpful suggestions.

Key Words: Screening, Multistage Selection, Information Overload, Bounded Rationality.

JEL Classification: D70, D80.

Addresses for Correspondence: Vladimir Karamychev, Department of Economics, Erasmus University Rotterdam, H07-19, Burg. Oudlaan 50, 3062 PA Rotterdam, The Netherlands; tel.: +31-(10)-408.2158, E-mail: karamychev@few.eur.nl. Stefano Ficco, Department of Economics and Tinbergen Institute, Erasmus University Rotterdam, H16-20, Burg. Oudlaan 50, 3062 PA Rotterdam, The Netherlands; tel.: +31-(10)-4088938. E-mail: ficco@few.eur.nl.

1. Introduction

Let us consider the following two examples. 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. There are two types of complications that make the decision-making in this example very difficult. The first comes form the consideration that companies have limited budgets and, consequently, the resources each department can allocate to projects' evaluations 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.

In the second example we look at an individual who must select a car for a purchase. He, first, looks through a car magazine and, based on the information he gathers there (i.e., prices, features, driving performance, etc.) selects a number of car models he may be interested in. Then, he goes to a car dealer where he personally compares the "look-and feel" of the cars selected form the magazine and, eventually, buys one of them. Like the budget of a company, the time and attention that a person can allocate to interpret different alternatives are limited. In our example one may think that the accuracy with which the individual is able to investigate each car decreases if his attention is spread over too many models. Moreover, as his decision must rely on different pieces of information obtained at different points in time, the information the individual remembers from the magazine can be crucial for the overall quality of the decision made.

These two examples clearly show that internal information-processing limitations may prevent an otherwise perfect decision-maker (an organization in the former example and an individual in the latter example) from selecting the best feasible alternative with certainty. It is the limited amount of information an individual can absorb at a time, and his limited memory capacity that are responsible for such imperfections. Similarly, organizations of individuals have constraints both in acquiring and communicating information. Hence, the amount of information may adversely affect the outcome of the selection procedure. Quantitatively, more information means more alternatives and this requires more resources to evaluate all alternatives. Qualitatively, more complex information requires more resources to evaluate each alternative.

In accordance with this dichotomy we distinguish two limitations in information processing. 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. We will refer to this limitation as *imperfect information acquisition*. The second limitation plays a role on the information transmission level due to imperfect communication between members of the same organization, or to the fact that an individual does not remember some information that he previously knew. Since a decision-maker can be thought as an information–processing network, this limitation can be interpreted as imperfect transmission form one node of the decision-maker's internal structure to another. We will, therefore, call it *imperfect information transmission*.

The aim of this paper is to incorporate both these information-processing limitations in a model of fully rational individuals. More precisely, we 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), or as one individual who moves at two different points in time (possibly, with imperfect recall), 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, i.e., high or low, signal about its quality. Having observed the signal outcomes, selector 1 selects a sub-sample of the 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 first 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 otherwise. Based on these observations, the R&D department selects a sub-sample of projects and passes it to the second selection stage, i.e., to the marketing department. Similarly, the latter applies its criterion to each alternative in the sub-sample and observes a high signal if a project meets the requirement.

In general settings, each type of alternatives may generate either a high or a low signal in every stage. We also consider two special cases of the model, where results can be generalized to an arbitrary number of stages. In the first case the screening requirements are set so high that no low quality alternative can ever meet them. We call this case *high-standard filtering selection* as only high types can pass screening filters by generating high signals. In the opposite case the screening requirements are set so low that every high quality alternative meets them for sure. We call this case *low-standard filtering selection* as only low types may fail to pass screening filters by generating low signals.

In our analysis we consider four cases. The first one is a benchmark case where the two information-processing limitations are absent: both transmission and acquisition of information are perfect. This situation is captured by assuming that all information obtained at stage 1, i.e., 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. We refer to this case as *full memory* and *constant accuracy* case (FM-CA case). The benchmark scenario is reminiscent of a perfect statistical environment where handling large samples is not costly and all the information gathered form sequential experiments could be used. In order to analyze the effect of each of the information-processing limitations on the behavior of rational agents, we then depart form the benchmark FM-CA scenario by analyzing three other cases. In the first one the information acquisition is kept perfect, but the assumption of perfect information transmission is relaxed by assuming that no information gathered in stage 1 is available in stage 2. We refer to this case as *no memory* and *constant accuracy* case (NM-CA case). In the second scenario, on the contrary, the information transmission is kept perfect, but not the information acquisition. We 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. We refer to this case as *full memory* and *decreasing accuracy* case (FM-DA case). Lastly, combining both types of imperfections we analyze the fourth scenario, which we refer to as *no memory* and *decreasing accuracy* case (NM-DA case).

It turns out that in any scenario, regardless of the underlying informational assumptions, there always multiple Nash equilibria exist. That is why we treat the problem of selecting the best alternative from a game-theoretic perspective rather than from a purely statistical point of view. However, there exists a unique trembling-hand perfect Bayes-Nash equilibrium (PBNE) for all generic values of the model's primitives.

The results are as follows. The PBNE in FM-CA case is such that independently of the signal outcomes in stage 1 all alternatives are passed to stage 2, and 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 the alternative with the largest likelihood value.

When we depart form 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 of by neglecting some potentially valuable information. We call this phenomenon *information overload*. The causes of information overload are different depending on the specific information-processing limitations.

Introducing the no memory assumption, 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 PBNE in NM-CA 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, in no-memory settings, information overload takes the form of 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.

In the FM-DA case, it is the decrease in accuracy in stage 2 that prevents selector 1 form taking too many alternatives, and that gives rise to information overload. The PBNE in FM-DA 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, in decreasing accuracy settings, information overload takes the form of a bound on the total number of alternatives, i.e., good and bad alternatives, to be passed to the next selection round. Contrary to the NM-CA case, in the FM-DA 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 negatively and monotonically affect it.

In the fourth scenario, when both sources of informational imperfections are present, information overload takes its most severe form as both imperfections work hand-in-hand and bound the sample size even further.

Finally, in multi-stage *filtering selection* we obtain the following results. In case of *high-standard filtering selection* information overload never occurs. 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. On the other hand, if the standards are *low*, only high signals are passed to the next stage, and on the later stages of the selection information will be heavily overloaded. More precisely, only two high signals are passed to the next stage provided that the alternatives have passed a sufficiently large number of filtering stages.

The rest of the paper is organized as follows. Section 2 reviews the existing related literature. Section 3 states the model. Section 4 analyzes four different informational scenarios, namely the benchmark FM-CA case and NM-CA, FM-DA and NM-DA cases. Section 5 analyzes two special filtering selection cases and section 6 concludes. The appendix contains all the proofs.

2. Related literature

This paper relates primarily to the bounded rationality literature on limited capacity. Since the assumption of perfect rationality has been questioned in the 50's, economic theory started a hunt for «Rational choice that takes into account the cognitive limitations of the decision maker – limitations of both knowledge and computational capacity», Simon (1987). In the existing literature much attention has been focused on finding plausible ways to model bounded rationality. Limited capacity models emphasize the role of cognitive heuristics and simplifying knowledge structures in reducing information-processing demands. Lipman (1995) provides an exhaustive survey of such economic literature where authors use different ways to model information-processing limitations.

Morris (1992), Lipman (1992) and Gilboa and Schmeidler (1992) follow an axiomatic approach. In non axiomatic models, processing limitations take a form of computational constraints like in Spear (1989), Anderlini (1991), and Anderlini and Felli (1992), or a form of a costly processing phase like in Rubinstein (1986), Abreu and Rubinstein (1988), and Kalai and Stanford (1988). Finally, Mount and Reiter (1990), and Radner and van Zandt (1992) study information processing limitations in the context of optimal processing networks.

Yet, there are relatively few papers that explore the implications of bounded rationality. The main reason is the lack of agreements on how to model this phenomenon. Our paper is a contribution in this direction as it enables us to *endogenously* explain a phenomenon of information overload, which is usually imposed *exogenously* in most bounded rationality models. Therefore, our model provides a rational foundation to bounded rationality in a form of information overload.

The idea that human brain is better equipped for working with relatively small number of alternatives has also been extensively exploited in the field of psychology. Baddley (1994) gives a good overview of this literature. More precisely, in his seminal paper, G. A. Miller (1956) pointed out that it is the span of absolute judgment and the span of immediate memory that imposes severe limitations on the amount of information we are able to simultaneously process and remember. We account for those two limitations in an economic framework by introducing imperfect information acquisition and imperfect information transmission.

Iimperfect information acquisition has also been present in many economic studies, both empirical and theoretical. Early, as well as more recent, works in marketing (see Jacoby et al. (1974) and Hahn et al. (1992), among others) provide evidence that supports the adverse effect of excessive information on the quality of the decision. Chernev (2003) points out that such phenomenon becomes even stronger when preferences do not have articulated attributes, and vanishes when preferences are more articulated. Most theoretical models (see van Zandt (2001), among others) introduce imperfect information acquisition in a form of an exogenous hard limit on the number of items an individual can process. In contrast, in Ficco (2004) imperfect information acquisition is also modeled as a sample size - accuracy trade-off, but its implications are studied in the context of a monopsony market.

Imperfect information transmission is taken into account by the literature on the economics of organizations, which is also closely related to our model. The idea that individuals have limited capacities to process information suggests that organizations (groups of individuals) may be able to make better decisions than any single individual. Consequently, a substantial amount of such literature views individual agents as single nodes in a large information-processing network of the organization. Although information transmission imperfections appeared in many of these studies (See, e.g., Radner (1993), Sah and Stiglitz (1986), Visser

(2000), among many others.), they have not yet been treated in the context of imperfect information acquisition in a fully rational environment.

Our model in no memory settings also studies the implications of imperfect recall on a multistage selection problem. A substantial amount of literature has been devoted to address different issues concerning the imperfect recall assumption in multistage games (see. i.e., Battigalli (1997), Gilboa (1997), Grove and Halpern (1997), Halpern (1997), Lipman (1997), and Piccione and Rubinstein (1997a, 1997b)).

We share with Moscarini and Smith (2002) the interpretation of the amount of information as the number of drawn signal outcomes. Finally, in our multistage selection model learning occurs via Bayes updating. On the contrary, Borgers, Morales and Sarin (2004) consider general learning rules in environments in which little prior and feedback information is available to the decision maker.

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^L respectively. The expected share of θ^H within the population is denoted by α. The game lasts two stages. In stage 1 each alternative generates a binary signal $s_1 \in \{s_1^H, s_1^L\}$, i.e., either a high signal s_1^H or a low signal s_1^L, which is correlated with its true type with the following revealing probabilities:

$$\Pr\left(s_1^H \middle| \theta^H\right) = q_1^H \in (0,1),\ \Pr\left(s_1^L \middle| \theta^L\right) = q_1^L \in (0,1).$$

Having observed a signal composition (H_1,L_1), i.e., H_1 high signals and L_1 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 $s_2 \in \{s_2^H, s_2^L\}$ in accordance with the revealing probabilities $q_2^H \in (0,1)$ and $q_2^L \in (0,1)$. In FM-case selector 2 makes his choice based on two signals, s_1 and s_2, observed in both stages. The pair (s_1, s_2) determines the overall likelihood value for each alternative. In NM-case, on the contrary, the signaling history is not available, and selector 2 makes his choice based only on signals s_2. Having observed realizations of signals in stage 2, selector 2 selects one alternative, which becomes the outcome of the selection procedure. We assume that all signals in stages 1 and 2 are statistically independent:

Assumption 1. $\Pr\left(s_1, s_2 \middle| \theta^i\right) = \Pr\left(s_1 \middle| \theta^i\right)\Pr\left(s_2 \middle| \theta^i\right)$, *i*=*H*, *L*.

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

Assumption 2. High and low signals are defined such that $q_t^H > 1 - q_t^L$, and, therefore, for all *t*:

$$\gamma_t^H \equiv \Pr\left(\theta^H \middle| s_t^H\right) = \frac{\alpha_t q_t^H}{\alpha_t q_t^H + (1-\alpha_t)\left(1-q_t^L\right)} > \frac{\alpha_t\left(1-q_t^H\right)}{\alpha_t\left(1-q_t^H\right) + (1-\alpha_t) q_t^L} = \Pr\left(\theta^H \middle| s_t^L\right) \equiv \gamma_t^L .$$

The information acquisition technology in the model is represented by revealing probabilities q_t^H and q_t^L. The numbers q_1^H and q_1^L, i.e., the screening accuracy in stage 1, are exogenously given. In contrast, it is the sample size $N_2 = h_1 + l_1$ that determines screening accuracy in stage 2. We assume that both q_2^H and q_2^L are either strictly decreasing functions of N_2 in case of DA, or they are constants in case of CA.

Since the revealing probability functions $q_2(n)$ are strictly decreasing and bounded functions in DA cases, they can be written as $q_2(n) = \underline{q} + f\left(\frac{1}{n}\right)$ where $\underline{q} \geq 0$ and $f(x)$ is a strictly increasing function such that $f(x)=0$. Function f can be treated as a production function of information acquisition technology and its argument $\frac{1}{n}$ 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 θ^H is eventually selected. An equilibrium strategy for selector 1 is the optimal sample composition $\left(h_1^*, l_1^*\right) = \left(h_1^*(H_1, L_1), l_1^*(H_1, L_1)\right)$ 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 Bayes-Nash equilibrium always exists, possibly in mixed strategies. Moreover, a pure strategy Bayes-Nash equilibrium always exists as the players are team members. In what follows we will consider Bayes-Nash equilibria in pure strategies only.

The model, regardless of the underlying informational assumptions, has always multiple Nash equlibria. In order to see why this is the case, consider the following strategy profile: selector 1 passes only one signal, preferably high, to stage2; selector 2 selects one alternative with the lowest likelihood of being θ^H type. Given the strategy of player 2, player 1 wants to effectively 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 who generated a high signal s_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 in what follows 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 FM-CA case, where the informational environment, apart from inaccurate screening, is perfect.

4.1.FM-CA case

In CA-case 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. This is the content of Proposition 1.

Proposition 1. In FM-CA case, the game has a unique PBNE such that:

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

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

i.e., player 1 selects all low signals if the screening accuracy in stage 2 is higher than in stage 1 and the other way around.

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

$$(s_1^H,s_2^H)\succ(s_1^L,s_2^H)\succ(s_1^H,s_2^L)\succ(s_1^L,s_2^L) \text{ if } \frac{q_2^H}{1-q_2^H}\frac{q_2^L}{1-q_2^L} > \frac{q_1^H}{1-q_1^H}\frac{q_1^L}{1-q_1^L},$$

$$(s_1^H,s_2^H)\succ(s_1^H,s_2^L)\succ(s_1^L,s_2^H)\succ(s_1^L,s_2^L) \text{ if } \frac{q_2^H}{1-q_2^H}\frac{q_2^L}{1-q_2^L} < \frac{q_1^H}{1-q_1^H}\frac{q_1^L}{1-q_1^L}.$$

Proposition 1 is proven as a sub-case of Proposition 3 in the appendix. It can be easily generalized to an arbitrary number of stages. Proposition 1 states one of the most general concepts in statistics that prescribes a calculation of full likelihoods for all available alternatives and then selection the maximum value. When the screening accuracy in stage 1 is higher than in stage 2, i.e., when $\Pr(\theta^H|s_1^H,s_2^L)>\Pr(\theta^H|s_1^L,s_2^H)$, and, therefore, $(s_1^H,s_2^L)\succ(s_1^L,s_2^H)$, only high signals are selected. If, on the contrary, stage 2 signaling is more accurate, i.e., $\Pr(\theta^H|s_1^H,s_2^L)<\Pr(\theta^H|s_1^L,s_2^H)$, selector 1 selects all the alternatives.

Having established the result in the benchmark case, we will see now how informational imperfections affect the resulting equilibrium strategies. First, we introduce imperfections in information transmission, which we classified as *no memory* case.

4.2.NM-CA case

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:

$$(s_2^H)\succ(s_2^L). \tag{1}$$

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 2 shows how the realization of signals in stage 1 affects the optimal sample composition.

Proposition 2. In NM-CA case there exists a generically unique PBNE 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 $\overline{L_1}(H_1)\in[0,\infty)$ and a lower-bound $\underline{L_1}(H_1)\in\left[0,\overline{L_1}(H_1)\right]$ such that:

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

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 $\overline{L_1}(H_1)$ are selected in stage 1.

c) $\overline{L_1}(H_1)$ does not increase and strictly decreases whenever $\overline{L_1}(H_1)>0$.

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

e) $l_1^*(h,L_1)=\underline{L_1}(h)=\overline{L_1}(h)=0$ for all $h \geq 1$ if $\frac{q_2^H}{1-q_2^H}\frac{q_2^L}{1-q_2^L}<\frac{q_1^H}{1-q_1^H}\frac{q_1^L}{1-q_1^L}$, 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, $\overline{L_1}(0)=+\infty$, $\underline{L_1}(0)=0$.

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 2 can be understood as follows. First, PBNE 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 memory 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.

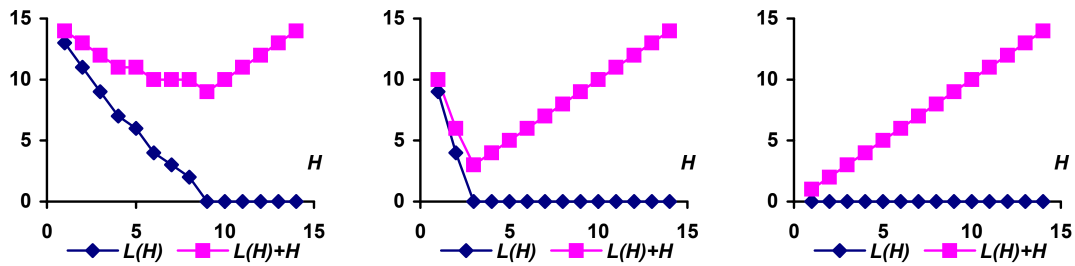
The sample size effect vanishes exponentially with the number of selected low signals while the mixing effect decreases reciprocally to that number. Hence, for a large number of selected low signals the latter dominates the former and there is an upper bound $\overline{L_1}$ such that $l_1^* \leq \overline{L_1}$. If only a few low signals are available, it might be possible that taking none of them is optimal even if $\overline{L_1} > 0$ 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 $\overline{L_1}$ 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 the no memory assumption leads to the overload of information in stage 1. This phenomenon comes into play when selector 1 observers heterogeneous signals. In this case 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, the rest will be neglected. Due to a non-monotone behavior of the total sample size $h_1^* + l_1^*$, it 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.

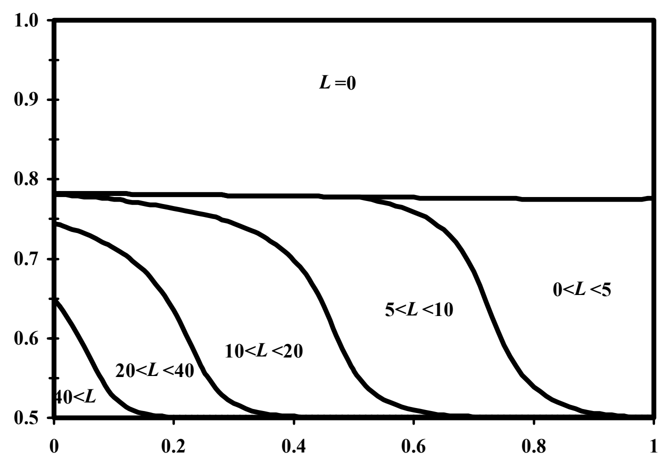
For exposition purposes we have numerically calculated the function $\overline{L_1}(H_1)$ in the following example.



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 $q_2^H = q_2^L = 0.9$.

Example 1.

Picture 1 shows the numerically calculated function $\overline{L_1}(H_1)$ for *α*=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 non-monotone behavior of the maximum sample size $\overline{L_1}(H_1) + H_1$.

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 $L = \overline{L_1}(1)$ takes different values, for the case $q_2^H = q_2^L = 0.9$. One may note that both the prior and the first stage accuracy monotonically and negatively affect the upper-bound $\overline{L_1}(1)$. This monotone dependence of $\overline{L_1}(H_1)$ is confirmed by numerous numerical calculations, yet the analytical proof is to be found. //

Summarizing, imperfect information transmission between selectors limits the number of low signals selected in stage 1. The mixing effect, which is responsible for such information overload, 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 NM-CA equilibrium we turn to another imperfection of information processing, namely to imperfect information acquisition. This is tFM-DA case.

4.3.FM-DA case

In FM-CA 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 $\left(s_1^H, s_2^L\right)$ is preferred to $\left(s_1^L, s_2^H\right)$ and the other way around. In FM-DA case, however, the accuracy in stage 2 is *endogenously* determined by the sample size, i.e., by the number of alternatives selected in stage 1. But 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 FM-CA case, selector 2 has a unique weakly undominated strategy, which is determined by its preference relation, which, in turn, is determined by selector 1. For the sake of the simplicity of exposition we assume here that both revealing probability functions coincide, i.e., $q_2^H(n) = q_2^L(n) = q_2(n)$.

It turns out that there exists a generically unique PBNE of the game, which always exhibits information overload in stage 1.

Proposition 3. In FM-DA case there exists a generically unique PBNE such that:

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

b) for any $H_1 \ge 0$ there exist an upper-bound $\overline{L_1}(H_1) \in [0,\infty)$ and a finite set of lower-bounds $\{\underline{L_1}^k\}_{k=1}^K$, $1 \le K < \overline{L_1}(H_1)$, $0 \le \underline{L_1}^1 < \underline{L_1}^k < \underline{L_1}^{k+1} \le \overline{L_1}(H_1)$ such that:

$$l_1^*(H_1,L_1) = \begin{cases} \min\{L_1, \overline{L_1}(H_1)\}, \text{if } L_1 \ge \underline{L_1}^K(H_1) \\ \underline{L_1}^k, \text{if } L_1 \in [\underline{L_1}^k, \underline{L_1}^{k+1}) \end{cases},$$

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 $\overline{L_1}(H_1)$ 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) $\overline{L_1}(H_1) = 0$ whenever $H_1 > \overline{H_1}$.

e) $l_1^*(h,L_1) = \underline{L_1}^1(h) = \overline{L_1}(h) = 0$ 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) Player 2 selects an alternative in accordance with the following preference relation:
$(s_1^H, s_2^H) \succ (s_1^L, s_2^H) \succ (s_1^H, s_2^L) \succ (s_1^L, s_2^L)$ if $q_2(H_2 + L_2) > q_1$,
$(s_1^H, s_2^H) \succ (s_1^H, s_2^L) \succ (s_1^L, s_2^H) \succ (s_1^L, s_2^L)$ if $q_2(H_2 + L_2) < q_1$.

The statements of Proposition 3 can be understood as follows. First, like in Proposition 2, PBNE 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 pay-offs. The first effect is the same *sample size* effect as in Proposition 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 2, prevents selecting too many signals of *both* types, thus, there are upper-bounds $\overline{L_1}$ and $\overline{H_1}$, 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 $\overline{L_1} > 0$. 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).

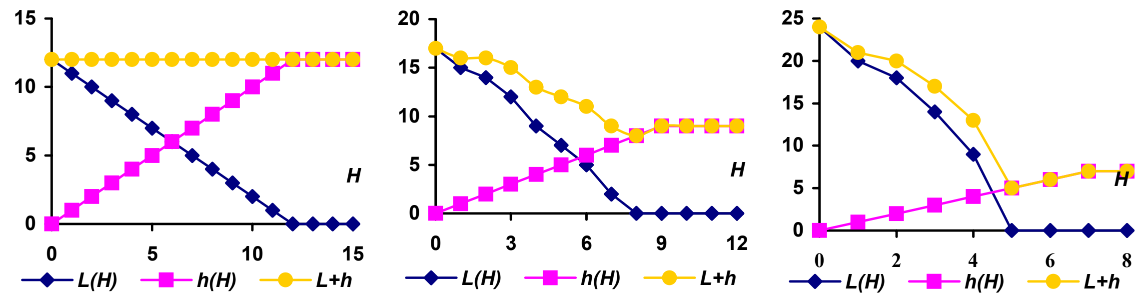
Comparing Proposition 2 and Proposition 3 one may note that the only difference between NM case and DA case is when signals in stage 1 are homogeneous: DA case exhibits information overload while NM case does not. In order show other differences between the two cases we provide numerically calculated functions $\overline{L_1}(H_1)$ and $h_1^*(H_1)$ in the following example.

Example 2.

In this example we have selected $q_2(n) = 0.89 + 0.01 \cdot \frac{2}{n}$ 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 generate the overload.

Picture 3 shows, first, that as in NM-CA case, the upper-bound $\overline{L_1}(H_1)$ is a non-increasing function and is strictly decreasing function when $\overline{L_1}(H_1) > 0$. Second, the maximum sample size in stage 1, just like in NM-CA case, is not a monotone function.

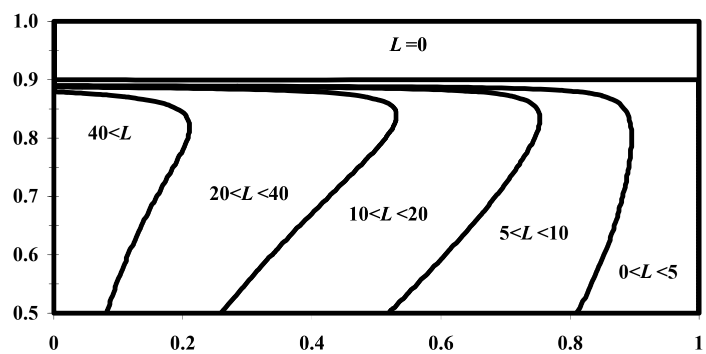
Like in NM-case, the prior α monotonically affects the sample composition: the higher the prior is, the less low signals are selected, see next Picture 4. In contrast, the accuracy in stage 1 affects the sample composition non-monotonically: the minimum low signals are selected either for very informative stage 1 signals or for almost uninformative ones. Contrary to the NM-CA case, under FM-DA 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 see, both types of information processing imperfections yield information overload in stage 1. The causes of the overload, however, are different. In NM case it is the purely statistical mixing effect that reduces the incentives to mix heterogeneous signals into a single pool. In DA case it is the decrease in accuracy that makes selection of large samples costly. Example 2 shows that even a tiny decrease in accuracy results in the overload. In order to highlight the common features the distinctions in information overload due to these two effects we compare equilibrium properties obtained analytically in Proposition 2 and Proposition 3, and obtained numerically in Example 1 and Example 2.

First of all, the mixing effect manifests itself only in heterogeneous samples. That is why in NM case the overload does not arise if only high or only low signals are observed.

Furthermore, when the accuracy in stage 1 vanishes, so does the overload in NM case. 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 overload, vanishes. In this case the overload prevents selecting low signals in stage 1 whatsoever. Therefore, and this is the second principal difference between NM and DA overloads, the accuracy in stage 1 affects the number of selected low signals non-monotonically in DA case whereas in NM case this dependence in monotone.

Apart from these two distinctions all the other equilibrium features of NM and DA scenarios 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 selected low signals 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 the high signals is sufficiently large, none of low signals will be selected.

Both effects become stronger also for large value of the prior. In other words, the overload 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 overload in NM and DA cases. The last, but not least, common feature of NM and DA 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.NM-DA case

We finish the analysis of the general 2-stage binary selection model by allowing both sources of imperfections, which is our NM-DA case. Combining Proposition 2 (NM case) and Proposition 3 (DA case) we get the following result.

Proposition 4. In NM-DA case there exists a generically unique PBNE such that:

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

b) for any $H_1 \geq 0$ there exist an upper-bound $\overline{L_1}(H_1) \in [0,\infty)$ and a finite set of lower-bounds $\{\underline{L_1}^k\}_{k=1}^K$, $1 \leq K < \overline{L_1}(H_1)$, $0 \leq \underline{L_1}^1 < \underline{L_1}^k < \underline{L_1}^{k+1} \leq \overline{L_1}(H_1)$ such that:

$$l_1^*(H_1, L_1) = \begin{cases} \min\{L_1, \overline{L_1}(H_1)\}, \text{if } L_1 \geq \underline{L_1}^K(H_1) \\ \underline{L_1}^k, \text{if } L_1 \in [\underline{L_1}^k, \underline{L_1}^{k+1}) \end{cases},$$

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 $\overline{L_1}(H_1)$ are selected in stage 1.

c) $\overline{L_1}(H_1)$ does not increase and strictly decreases whenever $\overline{L_1}(H_1) > 0$.

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

e) $l_1^*(h, L_1) = \underline{L_1}^1(h) = \overline{L_1}(h) = 0$ 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) 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 $l_1^* = 0$.

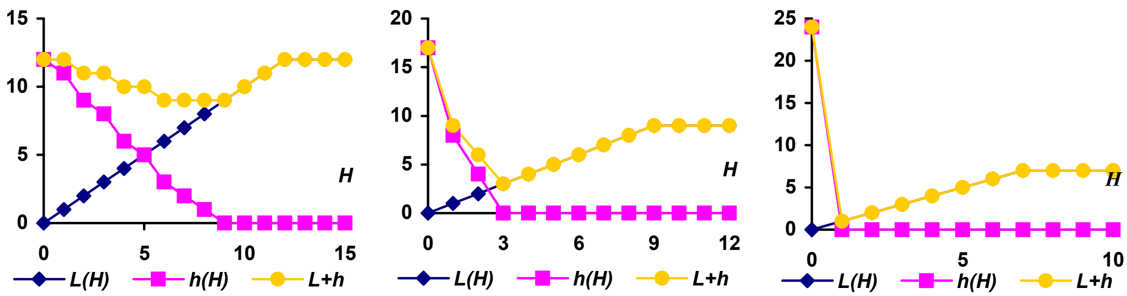
g) Player 2 selects an alternative in accordance with the preference relation (1).

The proof of Proposition 4 can be obtained by adjusting the proof of Proposition 2 to decreasing $q_2(n)$ function and, therefore, is omitted. Naturally, when both types of informational imperfections are present in the model, the overload of information is the largest. The following Example 3 shows the result of imposing the no memory assumption on the FM-DA case.

Example 3.

This example differs from Example 2 only in the memory assumption. Picture 5 shows that in NM-DA case the sample size in stage 2 is even smaller than in FM-DA 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.

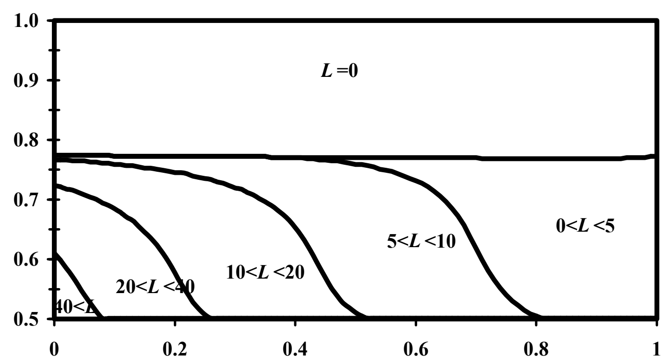
One may see that there hardly can be found any criteria that would allow us to classify both types of imperfections based on exogenous variables only. As we have already



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 $q_2(n) = 0.89 + 0.01 \cdot \frac{2}{n}$.

noticed, the major difference between NM and DA imperfections is in the way they react to changes in the accuracy of stage 1 screening, compare Picture 2, Picture 4 and Picture 6 //

We have seen that whatever informational imperfections are, they overload stage 1 selection. If it is imperfect information transmission then only low signals will be neglected due to the mixing effect. If it is imperfect information acquisition, 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. In NM settings 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 selected in stage 1 alternative requires accounting for a bi-variate binomial distribution of signal realizations in stage 2, a 4-variate distribution in stage 2 and, in general 2^{t-1}-variate distribution in stage *t*. In FM settings the very same distributions must be taken into account as well, though not within the beliefs on later stages but rather in calculating expected pay-offs in stage 1

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

We have started the paper by arguing that the decision-maker has internal informational imperfections that prevent him from achieving the first best outcome, which is the selection of the best feasible alternative with probability 1. In these circumstances it is very reasonable to assume that the very same decision-maker will have difficulties with selecting the optimal selection rule for every selection stage (PBNE equilibrium of the game) as well. Let us consider the problem of selection of the best feasible alternative in a wider framework.

A decision-maker (either a human or a firm) faces the following task. First, he has to decide how many selection stages he is going to implement. Second, he has to work out screening procedures for every stage. The latter means that the decision-maker derives imprecise performance measures, i.e., likelihood functions that will be computed for all previously selected alternatives at every stage. Third, the decision-maker has to find an optimal selection rule, i.e., PBNE of the selection model he has built. And finally fourth, he follows the rule in order to select the best alternative.

In the previous section we analyzed the selection game for arbitrary screening procedures on both stages. It turns out, that imposing either very strong or very weak requirements on all stages of the game allows us to generalize the model to an arbitrary number of selection stages.

When the passing requirements are so strong that only high type alternatives are able to meet them, i.e., to pass the filter, low type alternatives always generate low signals. In other words, $q_t^L = 1$. Thus, the screening *filters* all high signals as they have necessarily come from high types. When the passing requirements are so weak that only low type alternatives may fail to meet them, high type alternatives always generate high signals. In other words, $q_t^H = 1$. Thus, the screening *filters* all low signals out as they have necessarily come from low types.

We will call such selection procedures as *filtering selection*. In this section we will investigate properties of such filtering selection procedures. We begin with the high standard filter $q_t^L = 1$.

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 PBNE 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 subgame with an arbitrary number of stages. Indeed, without mixing effect, memory plays no role as every player, having observed more than one alternative, infers that all of them generated only low signals in the past.

It turns out that the high standard filtering selection procedure exhibits no information overload provided the initial set of alternatives is large enough.

Proposition 5. There exists a threshold level of the population size $\overline{N}$ such that for all $N > \overline{N}$ the filtering selection game with $q_t^L = 1$ has a unique PBNE $\{(h_t^*, l_t^*)\}_{t=1}^T$ such that for all $t = 1, \ldots, T$:

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

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

The proof of Proposition 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

Like in the previous case, mixing of types does not occur here. Indeed, 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 are pay-off equivalent and generate zero pay-off. Thus, the uniqueness property of PBNE fails. In what follows we assume that the player who observes only low signals is forced to select only *one* alternative, effectively ending the selection.

Contrary to the filter $q_t^L = 1$, filter $q_t^H = 1$ always exhibits information overload in a two-stage filtering selection game.

Proposition 6. In the filtering selection game with $q_1^H = q_2^H = 1$ there exists a generically unique PBNE such that:

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

b) $l_1^*(H_1, L_1) = 1 - sign(H_1)$, i.e., none of low signals are selected unless there are no high signal 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 6 can be easily obtained from the proof of Proposition 3 by taking $q_1^H = q_2^H = 1$ and, therefore, omitted. More interesting results, however, can be obtained for the case when the initial *prior* α is close enough to 1. The following proposition states the result.

Proposition 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 PBNE $\{(h_t^*, l_t^*)\}_{t=1}^T$ such that for all $t = 1, \ldots, T$:

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

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

We have seen in section 4 that in the general two-stage selection model the overload increases when the prior gets larger. Proposition 7 shows to what extent the overload limits the number of selected high signals: 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 $q_t^L(n)$. 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 7.

Corollary 1. If the number of selection stages T in the filtering selection game with the filter $q_t^H = 1$ is sufficiently large, then 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 conditions of Proposition 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 $n \geq \bar{n}$:

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

then player 1 never selects more than $\bar{n}$ high signals.[1] One can easily see that condition (2) is satisfied for $n \geq 2$ for any strictly decreasing $q_t^L(n)$ when α approaches 1, which is exploited in the proof of Proposition 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 $q_t^L(n) = 0.9 + 0.000001 \cdot \frac{1}{n}$ and for α=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.

[1] See appendix for the proof.

6. Conclusion

We developed a model of fully rational agents with internal informational imperfections. Those imperfections are introduced by assuming limits on information acquisition and information transmission. These assumptions are justified and supported by the clear analogy they have with the human brain's computational limitations already pointed out by experimental studies in psychology. With this framework we are able to obtain and explain information overload phenomenon. It enables us to challenge the paradigm «more information is better» as in our model neglecting valuable information emerges as an endogenous behavior of fully rational agents, while in most models of bounded rationality such behavior is exogenously imposed. The forces and mechanisms responsible for the overload are also investigated in deep details.

Even more striking results are obtained, when we turn our attention on screening procedures that take a form of filtering. When selection requirements at all stages are weak in a sense that good alternatives always satisfy them and bad alternatives are gradually filtered out, information overload appears in its most severe form: relatively few alternatives are sufficient in order to make an efficient choice. An 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 overload does not arise at all and the whole set of alternatives has to be passed to the next selection stage.

References

Abreu, D. and A. Rubinstein (1988). The Structure of Nash Equilibrium in Repeated Games with Finite Automata. *Econometrica*, 56, 1259-1282.

Anderlini, L. (1991). Communication, Computability, and Common Interest Games. Working paper, St. John's College, Cambrige.

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

Baddley, A. (1994). The Magical Number Seven: Still Magic After All These Years? *Psychological Review*, 101, 353-356.

Battigalli, P. (1997). Dynamic Consistency and Imperfect Recall. *Games and Economic Behavior*, 20, 31-50.

Borgers, T., A. J. Morales and R. Sarin (2004). Expedient and Monotone Learning Rules. *Econometrica*, 72, 383-405.

Chernev, A. (2003). Product Assortment and Individual Decision Processes. *Journal of Personality and Social Psychology*, 85, 151-162.

Ficco, S., (2004). Information Overload in Monopsony Markets. Tinbergen Institute Discussion Paper series.

Gilboa, I. (1997). A Comment on the Absent Minded Driver Paradox. *Games and Economic Behavior*, 20, 25-30.

Gilboa, I. and D. Schmeidler (1992). Additive Representations of Non-Additive Measures and The Choquet Integral. Working paper, Northwestern University.

Grove, A. and J. Halpern (1997). On the Expected Value of Games with Absentmindedness. *Games and Economic Behavior*, 20, 51-65.

Hahn, M., R. Lawson and G. Young (1992). The Effect of Time Pressure and Information Load on Decision Quality. *Psychology and Marketing*, 9, 365-379.

Halpern, J. (1997). On Ambiguities in the Interpretation of Games Trees. *Games and Economic Behavior*, 20, 66-96.

Jacoby, J., D. E. Speller and C. K. Berning (1974). Brand Choice Behavior as a Function of Information Load: Replication and Extension. *The Journal of Consumer Research*, 1, 33-42.

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

Lipman, B. L. (1992). Decision Theory with Impossible Possible Words. Working paper, Queen's University.

Lipman, B. L. (1995). Information Processing and Bounded Rationality: A Survay. *The Canandian Journal of Economics*, 28, 42-67.

Lipman, B. L. (1997). More Absentmindedness. *Games and Economic Behavior*, 20, 97-101.

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

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.

Morris, S. (1992). Revising Knowledge: A Decision-Theoretic Approach. Working paper, University of Pennsylvania.

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

Mount, K. and S. Reiter (1990). A Model of Computing with Human Agents. Working paper, Northwestern University.

Piccione, M. and A. Rubinstein (1997a). On the Interpretation of Decision Problems with Imperfect Recall. *Games and Economic Behavior*, 20, 3-24.

Piccione, M. and A. Rubinstein (1997b). The Absent Minded Driver's Paradox: Synthesis and Resposes. *Games and Economic Behavior*, 20, 121-130.

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

Radner, R. and T. van Zandt (1992). Information Processing and Returns to Scale. *Annales d'Economie et de Statistique*, 25/26, 265-298.

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

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

Simon, H. A. (1987). Bounded Rationality. In the New Palgrave, ed. J. Eatwell, M. Milgate, and P. Newman (New York: W.W. Norton).

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

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

Zandt, T., van, (2001). Information Overload in a Network of Targeted Communication, 2 May 2001, available at http://zandtwerk.insead.fr/papers/Overload.pdf.

Appendix

Proof of Proposition 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 turns out to be a rational analytical function of the model primitives α, 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 PBNE.

Then, we show that u 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 as a function of discrete argument L_1. It turns out that u 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 attains its global maximum:

$$\overline{L_1}(H_1)=\arg\max_{0\leq l_1} u(H_1,l_1).$$

If the interior maximum does not exist we define $\overline{L_1}(H_1)=0$ as in this case u strictly decreases for all l_1. Then, we define a lower-bound as the smallest l_1 that yields at least $u(H_1,0)$:

$$\underline{L_1}(H_1)=\begin{cases}0, \text{ if } u(H_1,l_1)<u(H_1,0) \text{ for all } l_1>0\\ \min\left\{l_1 \middle| 0<l_1\leq \overline{L_1}(H_1), u(H_1,l_1)\geq u(H_1,0)\right\} \text{ otherwise}\end{cases}$$

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 $\min\{L_1,\overline{L_1}(H_1)\}$ 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 functions $\overline{L_1}(H_1)$ 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 $l_1^*(h,L_1)=\underline{L_1}(h)=\overline{L_1}(h)=0$. Part (h) is trivial.

In what follows we use the following notations:

$$\varphi^{LL}=\gamma_1^L(1-q_2^H)+(1-\gamma_1^L)q_2^L,\ \varphi^{HL}=\gamma_1^H(1-q_2^H)+(1-\gamma_1^H)q_2^L,$$

$$\gamma^{HH}=\frac{\gamma_1^H q_2^H}{1-\varphi^{HL}},\ \gamma^{LH}=\frac{\gamma_1^L q_2^H}{1-\varphi^{LL}},\ \gamma^{HL}=\frac{\gamma_1^H(1-q_2^H)}{\varphi^{HL}},\ \gamma^{LL}=\frac{\gamma_1^L(1-q_2^H)}{\varphi^{LL}}.$$

Suppose player 1 selects h_1 signals s_1^H and l_1 signals s_1^L. 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 s_2^H. The probability of this event is given by the following bivariate binomial distribution:

$$\Pr(x, y|h_1, l_1) = \Pr(x|h_1)\cdot \Pr(y|h_1) = \binom{x}{h_1}(1-\varphi^{HL})^x (\varphi^{HL})^{h_1 - x} \cdot \binom{y}{l_1}(1-\varphi^{LL})^y (\varphi^{LL})^{l_1 - y}.$$

When this event occurs receiver 2 observes $x+y$ high signals s_2^H. 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:

$$u(h_1, l_1) = \sum_{\substack{x=y=0\\ x+y>0}}^{\substack{x=h_1\\ y=l_1}} \Pr(x, y|h_1, l_1)\left(\tfrac{x}{x+y}\gamma^{HH} + \tfrac{y}{x+y}\gamma^{LH}\right) + \Pr(0,0|h_1, l_1)\left(\tfrac{h_1}{h_1+l_1}\gamma^{HL} + \tfrac{l_1}{h_1+l_1}\gamma^{LL}\right)$$

$$= \gamma^{LH} + \left(\gamma^{HH} - \gamma^{LH}\right)\sum_{\substack{x=1\\ y=0}}^{\substack{x=h_1\\ y=l_1}} \Pr(x, y|h_1, l_1)\tfrac{x}{x+y} + \Pr(0,0|h_1, l_1)\left(\tfrac{h_1}{h_1+l_1}\left(\gamma^{HL} - \gamma^{LL}\right) - \left(\gamma^{LH} - \gamma^{LL}\right)\right).$$

Converting the finite sum above into an integral yields:

$$u(h_1, l_1) = \gamma^{LH} + \left(\gamma^{HH} - \gamma^{LH}\right)\int_{w=0}^{w=1}\left(\varphi^{LL} + (1-\varphi^{LL})w\right)^{l_1} d\left(\varphi^{HL} + (1-\varphi^{HL})w\right)^{h_1} +$$

$$+ \left(\varphi^{HL}\right)^{h_1}\left(\varphi^{LL}\right)^{l_1}\left(\tfrac{h_1}{h_1+l_1}\left(\gamma^{HL} - \gamma^{LL}\right) - \left(\gamma^{LH} - \gamma^{LL}\right)\right).$$

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

$$u(h_1+1, l_1) = \left(1-\varphi^{HL}\right)\sum_{x=y=0}^{\substack{x=h_1\\ y=l_1}} \Pr(x, y|h_1, l_1)\left(\tfrac{x+1}{x+1+y}\gamma^{HH} + \tfrac{y}{x+1+y}\gamma^{LH}\right) +$$

$$+ \varphi^{HL}\left(\sum_{\substack{x=y=0\\ x+y>0}}^{\substack{x=h_1\\ y=l_1}} \Pr(x, y|h_1, l_1)\left(\tfrac{x}{x+y}\gamma^{HH} + \tfrac{y}{x+y}\gamma^{LH}\right) + \Pr(0,0|h_1, l_1)\left(\tfrac{h_1+1}{h_1+1+l_1}\gamma^{HL} + \tfrac{l_1}{h_1+1+l_1}\gamma^{LL}\right)\right),$$

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

$$\Delta_h u(h_1, l_1) = \Pr(0,0|h_1, l_1)\left(\left(1-\varphi^{HL}\right)\left(\gamma^{HH} - \gamma^{HL}\right) + \frac{l_1\left(\gamma^{HL} - \gamma^{LL}\right)}{(h_1+1+l_1)}\left(\left(1-\varphi^{HL}\right) + \frac{1}{(h_1+l_1)}\right)\right) +$$

$$+ \left(1-\varphi^{HL}\right)\left(\gamma^{HH} - \gamma^{LH}\right)\sum_{x=y>0}^{\substack{x=h_1\\ y=l_1}} \Pr(x, y|h_1, l_1)\frac{y}{(x+1+y)(x+y)}$$

$$> 0.$$

Thus, $h_1^*(H_1, L_1) = H_1$.

Let us now fix any $h_1 \geq 1$ and define

$$D(h_1,l_1) \equiv \frac{\Delta_l u(h_1,l_1)}{h_1(\varphi^{HL})^{h_1}(\varphi^{LL})^{l_1}(\gamma^{HL}-\gamma^{LL})} \equiv \frac{u(h_1,l_1+1)-u(h_1,l_1)}{h_1(\varphi^{HL})^{h_1}(\varphi^{LL})^{l_1}(\gamma^{HL}-\gamma^{LL})},$$

$$D(h_1,l_1) = -\frac{\gamma^{HH}-\gamma^{LH}}{\gamma^{HL}-\gamma^{LL}}\frac{(1-\varphi^{HL})(1-\varphi^{LL})}{\varphi^{HL}}\int_{w=0}^{w=1}(1-w)\left(1+\frac{1-\varphi^{LL}}{\varphi^{LL}}w\right)^{l_1}\left(1+\frac{1-\varphi^{HL}}{\varphi^{HL}}w\right)^{h_1-1}dw + \\ +\frac{\varphi^{LL}}{h_1+1+l_1}-\frac{1}{h_1+l_1}+\frac{\gamma^{LH}-\gamma^{LL}}{\gamma^{HL}-\gamma^{LL}}\frac{(1-\varphi^{LL})}{h_1},$$

and

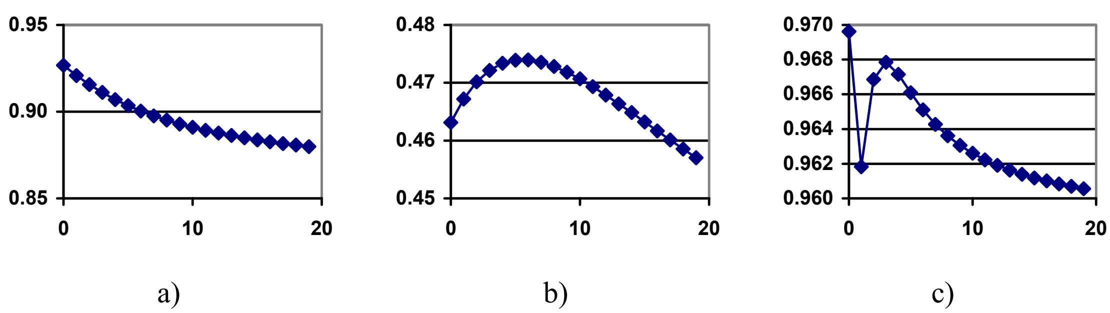
$$E(h_1,l_1) \equiv \Delta_l D(h_1,l_1) \equiv D(h_1,l_1+1)-D(h_1,l_1),$$

$$E(h_1,l_1) = -\frac{\gamma^{HH}-\gamma^{LH}}{\gamma^{HL}-\gamma^{LL}}\frac{(1-\varphi^{HL})(1-\varphi^{LL})^2}{\varphi^{HL}\varphi^{LL}}\int_{w=0}^{w=1}(1-w)w\left(1+\frac{1-\varphi^{LL}}{\varphi^{LL}}w\right)^{l_1}\left(1+\frac{1-\varphi^{HL}}{\varphi^{HL}}w\right)^{h_1-1}dw + \\ +\frac{1}{(h_1+1+l_1)(h_1+2+l_1)}\left((1-\varphi^{LL})+\frac{2}{(h_1+l_1)}\right).$$

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

a) $E(h_1,0)<0$. In this case $E(h_1,l_1)<E(h_1,0)<0$ and, therefore, $D(h_1,l_1+1)<D(h_1,l_1)$ and $\lim_{l_1\to\infty}D(h_1,l_1)=-\infty$. Thus, $D(h_1,0)$ determines the behavior of $u(h_1,l_1)$. If $D(h_1,0)<0$ then $D(h_1,l_1)<0$ 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, $D(h_1,0)>0$, then u first increases to its interior maximum $\overline{L_1}(H_1)$ 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) $E(h_1,0)>0$. In this case there exists a number X such that $E(h_1,X)<0<E(h_1,X-1)$, i.e., $D(h_1,l_1)$ has a unique maximum at $l_1=X$. If $D(h_1,X)<0$ then $D(h_1,l_1)<0$ and u always decreases, as in Picture 7(a). If, on the other hand, $D(h_1,X)>0$ then u has a unique



Picture 7.

interior maximum at $l_1 > X$ and decreases afterwards. In this case, if $D(h_1,0) > 0$ u increases for all $l_1 < X$, as in Picture 7(b); if $D(h_1,0) < 0$ 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).

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 $\overline{L_1}(H_1)$ 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, $l_1 = \overline{L_1}(H_1) > 0$ and $l_1 = \overline{L_1}(H_1 + 1) > 0$ respectively. We will show that $\overline{L_1}(H_1 + 1) \leq \overline{L_1}(H_1) - 1$. To this end we note that $D(H_1, l_1) < 0$ for all $l_1 \geq \overline{L_1}(H_1)$. Let us consider a difference $F \equiv D(H_1 + 1, l_1 - 1) - D(H_1, l_1)$:

$$F = -\frac{\gamma^{LH} - \gamma^{LL}}{\gamma^{HL} - \gamma^{LL}} \frac{(1 - \varphi^{LL})}{H_1(H_1 + 1)} - \frac{\gamma^{HH} - \gamma^{LH}}{\gamma^{HL} - \gamma^{LL}} \frac{(1 - \varphi^{HL})(1 - \varphi^{LL})(\varphi^{LL} - \varphi^{HL})}{(\varphi^{HL})^2 \varphi^{LL}} \times$$
$$\times \int_{w=0}^{w=1} (1 - w) w \left(1 + \frac{1 - \varphi^{LL}}{\varphi^{LL}} w\right)^{l_1 - 1} \left(1 + \frac{1 - \varphi^{HL}}{\varphi^{HL}} w\right)^{H_1 - 1} dw -$$
$$< 0,$$

as $\varphi^{LL} - \varphi^{HL} = (\gamma_1^H - \gamma_1^L)(q_2^H + q_2^L - 1) > 0$. Thus, $D(H_1 + 1, l_1) < 0$ for all $l_1 \geq \overline{L_1}(H_1) - 1$ and, therefore, $\overline{L_1}(H_1 + 1) \leq \overline{L_1}(H_1) - 1$, i.e., $\overline{L_1}(H_1)$ strictly decreases.

b) Suppose that for $h_1 = H_1 + 1$ $u(h_1, l_1)$ attains its global maximum at the corner $l_1 = \overline{L_1}(H_1 + 1) = 0$. Then trivially $0 = \overline{L_1}(H_1 + 1) \leq \overline{L_1}(H_1) \geq 0$.

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 $l_1 = \overline{L_1}(H_1) = 0$ while for $h_1 = H_1 + 1$ it attains its global maximum at the interior point $l_1 = \overline{L_1}(H_1 + 1) > 0$. We will show that this can never be true.

As $l_1 = \overline{L_1}(H_1 + 1)$ 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 $l_1 = \overline{L_1}(H_1 + 1)$.[2] In addition, as at $h_1 = H_1$ $l_1 = \overline{L_1}(H_1) = 0$ is assumed to be a global maximum, it must be that:

[2] This inequality is nothing more than the second order condition in the discrete form.

$$G(H_1,l_1) \equiv \frac{u(H_1,0)-u(H_1,l_1)}{(\varphi^{HL})^{H_1-1}(\gamma^{HL}-\gamma^{LL})}$$

$$= \frac{\gamma^{HH}-\gamma^{LH}}{\gamma^{HL}-\gamma^{LL}}(1-\varphi^{HL})H_1 \int_{w=0}^{w=1} \left(1-(\varphi^{LL}+(1-\varphi^{LL})w)^{l_1}\right)\left(1+\frac{1-\varphi^{HL}}{\varphi^{HL}}w\right)^{H_1-1} dw +$$

$$+\varphi^{HL}\left(\frac{\gamma^{HL}-\gamma^{LH}}{\gamma^{HL}-\gamma^{LL}}\left(1-(\varphi^{LL})^{l_1}\right)+\frac{l_1}{H_1+l_1}(\varphi^{LL})^{l_1}\right)$$

$$> 0$$

for all $l_1 > 0$. But then

$$Q \equiv \frac{(1-\varphi^{LL})}{l_1\varphi^{HL}(\varphi^{LL})^{l_1}}\Delta_h G(H_1,l_1) = \frac{(1-\varphi^{LL})}{l\varphi^{HL}(\varphi^{LL})^{l}}\left(G(H_1+1,l_1)-F(H_1,l_1)\right)$$

$$= -\frac{(1-\varphi^{LL})}{(H_1+l_1)(H_1+1+l_1)} + \frac{\gamma^{HH}-\gamma^{LH}}{\gamma^{HL}-\gamma^{LL}}\frac{(1-\varphi^{HL})(1-\varphi^{LL})}{l_1\varphi^{HL}(\varphi^{LL})^{l_1}} \times$$

$$\times \int_{w=0}^{w=1}\left(1-(\varphi^{LL}+(1-\varphi^{LL})w)^{l_1}\right)\left(1+(H_1+1)\frac{1-\varphi^{HL}}{\varphi^{HL}}w\right)\left(1+\frac{1-\varphi^{HL}}{\varphi^{HL}}w\right)^{H_1-1} dw$$

$$> -\frac{(1-\varphi^{LL})}{(H_1+l_1)(H_1+1+l_1)} + \frac{\gamma^{HH}-\gamma^{LH}}{\gamma^{HL}-\gamma^{LL}}\frac{(1-\varphi^{HL})(1-\varphi^{LL})^2}{\varphi^{HL}\varphi^{LL}} \times$$

$$\times \int_{w=0}^{w=1}(1-w)\left(1+\frac{1-\varphi^{LL}}{\varphi^{LL}}w\right)^{l_1-1}\left(1+\frac{H_1(1-\varphi^{HL})w}{\varphi^{HL}+(1-\varphi^{HL})w}\right)\left(1+\frac{1-\varphi^{HL}}{\varphi^{HL}}w\right)^{H_1} dw$$

$$> \frac{2\varphi^{LL}}{(H_1+l_1)(H_1+1+l_1)(H_1+2+l_1)} - E(H_1+1,l_1-1) + \frac{\gamma^{HH}-\gamma^{LH}}{\gamma^{HL}-\gamma^{LL}}\frac{(1-\varphi^{HL})(1-\varphi^{LL})^2}{\varphi^{HL}\varphi^{LL}} \times$$

$$\times \int_{w=0}^{w=1}(1-w)\left(1+\frac{1-\varphi^{LL}}{\varphi^{LL}}w\right)^{l_1-1}\left((1-w)+\frac{H_1(1-\varphi^{HL})w}{\varphi^{HL}+(1-\varphi^{HL})w}\right)\left(1+\frac{1-\varphi^{HL}}{\varphi^{HL}}w\right)^{H_1} dw$$

$$> 0$$

as $E(H_1+1,l_1-1)<0$ at $l_1=\overline{L_1}(H_1+1)$. 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 $l_1=\overline{L_1}(H_1+1)>0$.

All the three cases prove part (c) of the proposition. Part (d) follows from

$$u(0,l_1+1)-u(0,l_1) = (\gamma^{LH}-\gamma^{LL})(\varphi^{LL})^{l_1}(1-\varphi^{LL})>0.$$

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

Proof of Proposition 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 that proves Proposition 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 1.

As in the proof of Proposition 2, $u(h_1, l_1)$ turns out to be a rational analytical function of the model primitives $\alpha, q_1^H, q_1^L, \{q_2(n)\}_{n=1}^N$ and, therefore, it takes generically different values for different values its arguments $(h_1, l_1) \in [0, H_1] \times [0, L_1]$. Hence, there exists a generically unique PBNE. 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 $\overline{H_1}$ and $\overline{L_1}$, 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:

$$\varphi^{LL}(n) = \gamma_1^L(1 - q_2(n)) + (1 - \gamma_1^L)q_2(n),\ \varphi^{HL}(n) = \gamma_1^H(1 - q_2(n)) + (1 - \gamma_1^H)q_2(n),$$

$$\gamma^{HH}(n) = \frac{\gamma_1^H q_2(n)}{1 - \varphi^{HL}(n)},\ \gamma^{LH}(n) = \frac{\gamma_1^L q_2(n)}{1 - \varphi^{LL}(n)},\ \gamma^{HL}(n) = \frac{\gamma_1^H(1 - q_2(n))}{\varphi^{HL}(n)},\ \gamma^{LL}(n) = \frac{\gamma_1^L(1 - q_2(n))}{\varphi^{LL}(n)}.$$

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:

$$\begin{aligned} u(h_1, l_1) &= \left(1 - (\varphi^{HL})^{h_1}\right)\gamma^{HH} + (\varphi^{HL})^{h_1}\left(\left(1 - (\varphi^{LL})^{l_1}\right)\gamma^{LH} + (\varphi^{LL})^{l_1}\left(\gamma^{LL} + sign(h_1)(\gamma^{HL} - \gamma^{LL})\right)\right) \\ &= \gamma^{HH} - (\varphi^{HL})^{h_1}\left((\gamma^{HH} - \gamma^{LH}) + (\varphi^{LL})^{l_1}\left((\gamma^{LH} - \gamma^{LL}) - sign(h_1)(\gamma^{HL} - \gamma^{LL})\right)\right). \end{aligned}$$

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

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

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 $n \geq 1$ denotes the marginal disutility of having an extra high signal. Then, using (A.1) yields

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

Thus, for L_1=0 there exists an upper bound $\overline{H_1}<\infty$ such that $h_1^*(H_1,0)\le\overline{H_1}$.

Then, it is easy to see that $\Delta\equiv u(h+1,n-h-1)-u(h,n-h)>0$:

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

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 $h_1^*=H_1\ge 1$, let $MU(H_1,n-H_1)=u(H_1,n-H_1)-u(H_1,n-H_1+1)$ denotes the marginal disutility of having an extra low signal. Then we define:

$$F(H_1)\equiv\lim_{n\to\infty}\frac{MU(H_1,n-H_1)}{q_2(n)-q_2(n+1)},$$

$$F(H_1)=\frac{\gamma_1^H(1-\gamma_1^H)}{\left(1-\varphi^{HL}\right)^2}\left(1-\left(\varphi^{HL}\right)^{H_1}\right)+\frac{\gamma_1^L(1-\gamma_1^L)}{\left(1-\varphi^{LL}\right)^2}\left(\varphi^{HL}\right)^{H_1}+H_1\frac{\left(2\gamma_1^H-1\right)\left(\gamma_1^H-\gamma_1^L\right)}{\left(1-\varphi^{HL}\right)\left(1-\varphi^{LL}\right)}\left(1-\underline{q_2}\right)\underline{q_2}\left(\varphi^{HL}\right)^{H_1-1}$$

We will show that $F(H_1)>0$. It is clear that $F(H_1)>0$ for any $\gamma_1^H\ge\frac{1}{2}$. On order to show that $F(H_1)>0$ also for $\gamma_1^H<\frac{1}{2}$, we define

$$\begin{aligned}G(h)&\equiv\frac{\left(F(h+1)-F(h)\right)\left(1-\varphi^{LL}\right)^2\left(1-\varphi^{HL}\right)}{\left(\gamma_1^H-\gamma_1^L\right)\left(\varphi^{HL}\right)^{h-1}}\\ &=\left(\left(1-\left(\gamma_1^H+\gamma_1^L\right)\right)\left(1-\underline{q_2}\right)^2-\gamma_1^H\gamma_1^L\left(2\underline{q_2}-1\right)\right)\varphi^{HL}-\left(\varphi^{HL}-h\left(1-\varphi^{HL}\right)\right)\left(1-\varphi^{LL}\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q_2}\right)\underline{q_2},\end{aligned}$$

and consider 2 cases:

a) Let $\left(1-\left(\gamma_1^H+\gamma_1^L\right)\right)\left(1-\underline{q_2}\right)^2\ge\gamma_1^H\gamma_1^L\left(2\underline{q_2}-1\right)+\left(1-\varphi^{LL}\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q_2}\right)\underline{q_2}$. As G strictly increases, $G'(h)=\left(1-\varphi^{HL}\right)\left(1-\varphi^{LL}\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q_2}\right)\underline{q_2}>0$,

$$G(h)\ge G(0)=\left(\left(1-\left(\gamma_1^H+\gamma_1^L\right)\right)\left(1-\underline{q_2}\right)^2-\gamma_1^H\gamma_1^L\left(2\underline{q_2}-1\right)-\left(1-\varphi^{LL}\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q_2}\right)\underline{q_2}\right)\varphi^{HL}\ge 0.$$

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

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

b) Let $\left(1-\left(\gamma_1^H+\gamma_1^L\right)\right)\left(1-\underline{q_2}\right)^2<\gamma_1^H\gamma_1^L\left(2\underline{q_2}-1\right)+\left(1-\varphi^{LL}\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q_2}\right)\underline{q_2}$. In this case:

$$F(h)=\frac{\gamma_1^H\left(1-\gamma_1^H\right)}{\left(1-\varphi^{HL}\right)^2}-\left(\left(\frac{\gamma_1^H\left(1-\gamma_1^H\right)}{\left(1-\varphi^{HL}\right)^2}-\frac{\gamma_1^L\left(1-\gamma_1^L\right)}{\left(1-\varphi^{LL}\right)^2}\right)+h\frac{\left(1-2\gamma_1^H\right)\left(\gamma_1^H-\gamma_1^L\right)}{\left(1-\varphi^{HL}\right)\left(1-\varphi^{LL}\right)\varphi^{HL}}\left(1-\underline{q}_2\right)\underline{q}_2\right)\left(\varphi^{HL}\right)^h,$$

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

$$F(h)>\frac{\gamma_1^H\left(1-\gamma_1^H\right)\left(1-\varphi^{LL}\right)\varphi^{HL}-\left(\gamma_1^H-\gamma_1^L\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q}_2\right)\underline{q}_2\left(\varphi^{HL}+h\left(1-\varphi^{HL}\right)\right)\left(\varphi^{HL}\right)^h}{\left(1-\varphi^{HL}\right)^2\left(1-\varphi^{LL}\right)\varphi^{HL}}.$$

Let $Q(h)=\gamma_1^H\left(1-\gamma_1^H\right)\left(1-\varphi^{LL}\right)\varphi^{HL}-\left(\gamma_1^H-\gamma_1^L\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q}_2\right)\underline{q}_2\left(\varphi^{HL}+h\left(1-\varphi^{HL}\right)\right)\left(\varphi^{HL}\right)^h$

such that $F(h)>\frac{Q(h)}{\left(1-\varphi^{HL}\right)^2\left(1-\varphi^{LL}\right)\varphi^{HL}}$. Then

$$Q'(h)=-\left(\gamma_1^H-\gamma_1^L\right)\left(1-2\gamma_1^H\right)\left(1-\underline{q}_2\right)\underline{q}_2\left(\left(1-\varphi^{HL}\right)+\left(\varphi^{HL}+h\left(1-\varphi^{HL}\right)\right)\ln\varphi^{HL}\right)\left(\varphi^{HL}\right)^h.$$

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

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

Thus,

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

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 $\overline{L_1}(H_1)<\infty$ such that $l_1^*\leq\overline{L_1}(H_1)$.

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:

$$F(0)\equiv\lim_{n\to\infty}\frac{MU(0,n)}{q_2(n)-q_2(n+1)}=\lim_{n\to\infty}\frac{\gamma^{LH}(n)-\gamma^{LH}(n+1)}{q_2(n)-q_2(n+1)}=\frac{\gamma_1^L\left(1-\gamma_1^L\right)}{\left(1-\varphi^{LL}\right)^2}>0.$$

That proves the existence of an upper bound $\overline{L_1}(0)<\infty$.

The set of lower-bounds is recursively defined as follows:

$$\underline{L_1}^1 = \begin{cases} 0, \text{ if } u(H_1,l_1) < u(H_1,0) \text{ for all } l_1 > 0 \\ \min\left\{l_1 \middle| 0 < l_1 \le \overline{L_1}(H_1), u(H_1,l_1) \ge u(H_1,0)\right\} \text{ otherwise} \end{cases}$$

$$\underline{L_1}^{k+1} = \begin{cases} \underline{L_1}^k, \text{ if } u(H_1,l_1) < u\left(H_1,\underline{L_1}^k\right) \text{ for all } l_1 > \underline{L_1}^k \\ \min\left\{l_1 \middle| \underline{L_1}^k < l_1 \le \overline{L_1}(H_1), u(H_1,l_1) \ge u\left(H_1,\underline{L_1}^k\right)\right\} \text{ otherwise} \end{cases}$$

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 $l_1^*(H_1,L_1) = \arg\max_{0 \le l_1 \le L_1} u(H_1,l_1)$, is equal to $\underline{L_1}^k$ when $\underline{L_1}^k \le L_1 < \underline{L_1}^{k+1}$ and is equal to $\min\left\{L_1, \overline{L_1}(H_1)\right\}$ when $L_1 \ge \underline{L_1}^K$, that ends the proof of the proposition. ■

Proof of Proposition 5. In what follows we will use the following notations:

$$\varphi_t^L(\alpha_t,n_t) \equiv 1 - \alpha_t q_t^L(n_t),\; \gamma_t^L(\alpha_t,n_t) \equiv \frac{\alpha_t\left(1 - q_t^L(n_t)\right)}{\varphi_t^L(\alpha_t,n_t)}.$$

Here α_t stands for the *prior* in 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 s_t^H, the strategy $h_t^* = 1$ weakly dominates all the others. Thus, $h_t^* = \text{sign}(H_t)$. The rest of the proof is based on the induction assuming that only low signals were available.

The *ex-ante* pay-off function $\overline{u_T}(\alpha_T,n_T)$ in the last stage *T* is given by

$$\overline{u_T}(\alpha_T,n_T) = \left(1 - \left(\varphi_T^L(\alpha_T,n_T)\right)^{n_T}\right) \cdot 1 + \left(\varphi_T^L(\alpha_T,n_T)\right)^{n_T} \cdot \gamma_T^L$$

$$= 1 - (1-\alpha_T)\left(1 - \alpha_T q_T^L(n_T)\right)^{n_T-1} = 1 - (1-\alpha_T)\left(1 - \alpha_T\left(1 - \prod_{k=T}^{T}\left(1 - q_T^L(n_T)\right)\right)\right)^{n_T-1}.$$

Suppose that in stage *t* the *ex-ante* pay-off function $\overline{u_t}(\alpha_t,n_t)$ is

$$\overline{u_t}(\alpha_t,n_t) = 1 - (1-\alpha_t)\left(1 - \alpha_t q_t^L(n_t)\right)^{n_t-1} = 1 - (1-\alpha_t)\left(1 - \alpha_t\left(1 - \prod_{k=t}^{T}\left(1 - q_k^L(n_t)\right)\right)\right)^{n_t-1}.$$

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

$$u_{t-1}(\alpha_{t-1},l_{t-1}) = 1-\left(1-\gamma_{t-1}^{L}\right)\left(1-\gamma_{t-1}^{L}\left(1-\prod_{k=t}^{T}\left(1-q_{k}^{L}(l_{t-1})\right)\right)\right)^{l_{t-1}-1}$$

$$= 1-\frac{1-\alpha_{t-1}}{1-\alpha_{t-1}q_{t-1}^{L}(n_{t-1})}\left(1-\frac{\alpha_{t-1}\left(1-q_{t-1}^{L}(n_{t-1})\right)}{1-\alpha_{t-1}q_{t-1}^{L}(n_{t-1})}\left(1-\prod_{k=t}^{T}\left(1-q_{k}^{L}(l_{t-1})\right)\right)\right)^{l_{t-1}-1}.$$

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 $\overline{N_{t-1}}$ such that for all $n_{t-1}>\overline{N_{t-1}}$: $\arg\max_{l_{t-1}\le n_{t-1}} u_{t-1}=n_{t-1}$. 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 $l_{t}^{*}=\left(1-\mathrm{sign}(H_{t})\right)L_{t}$.

Then the *ex-ante* pay-off function $\overline{u_{t-1}}(\alpha_{t-1},n_{t-1})$ in stage t-1 becomes:

$$\overline{u_{t-1}}(\alpha_{t-1},n_{t-1}) = \left(1-\left(\varphi_{t-1}^{L}(\alpha_{t-1},n_{t-1})\right)^{n_{t-1}}\right)\cdot 1+\left(\varphi_{t-1}^{L}(\alpha_{t-1},n_{t-1})\right)^{n_{t-1}}\cdot u_{t-1}(\alpha_{t-1},l_{t-1}^{*})$$

$$= 1-\left(1-u_{t-1}(\alpha_{t-1},n_{t-1})\right)\left(\varphi_{t-1}^{L}(\alpha_{t-1},n_{t-1})\right)^{n_{t-1}}$$

$$= 1-\left(1-\alpha_{t-1}\right)\left(1-\alpha_{t-1}\left(1-\prod_{k=t-1}^{T}\left(1-q_{k}^{L}(n_{t-1})\right)\right)\right)^{n_{t-1}-1}.$$

Thus, for any $t=1,...,T$ there exists $\overline{N_{t}}$ such that $l_{t}^{*}=\left(1-\mathrm{sign}(H_{t})\right)L_{t}$ for all $n_{t}>\overline{N_{t}}$. Taking $\overline{N}=\overline{N_{1}}$ ends the proof. ■

Proof of Proposition 7. In what follows we will use the following notations:

$$\varphi_{t}^{L}(\alpha_{t},n_{t})\equiv(1-\alpha_{t})q_{t}^{L}(n_{t}),\ \gamma_{t}^{H}(\alpha_{t},n_{t})\equiv\frac{\alpha_{t}}{1-\varphi_{t}^{L}(\alpha_{t},n_{t})}.$$

Next, 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 $\overline{u_{T}}(\alpha_{T},n_{T})$ in the last stage, that defines the reduced form pay-off function in stage T-1, i.e., $u_{T-1}(\alpha_{T-1},h_{T-1})=\overline{u_{T}}(\alpha_{T}(\alpha_{T-1},h_{T-1}),h_{T-1})$ 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 $l_{T-1}^{*}=1-\mathrm{sign}(H_{T-1})$ for all $\alpha_{T-1}\in(\alpha_{T-1}^{*},1)$.

Next, we derive the *ex-ante* pay-off function $\overline{u_{T-1}}(\alpha_{T-1},n_{T-1})$ in stage T-1. We generalize it to an arbitrary stage t, i.e., $\overline{u_{t}}(\alpha_{t},n_{t})$ 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 $l_{t-1}^{*}=1-\mathrm{sign}(H_{t-1})$ 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 α_T, selector T selects a high signal, if there are, and gets a pay-off $\gamma_T^H(\alpha_T, n_T)$, which happens with probability $1-(\varphi_T^L(\alpha_T, n_T))^{n_T}$. With the remaining probability $(\varphi_T^L(\alpha_T, n_T))^{n_T}$ all signals in stage T are low and therefore, he selects one of them and gets zero. Thus

$$\overline{u_T}(\alpha_T, n_T) = \left(1-(\varphi_T^L(\alpha_T, n_T))^{n_T}\right)\gamma_T^H = \alpha_T \frac{1-(\varphi_T^L(\alpha_T, n_T))^{n_T}}{1-\varphi_T^L(\alpha_T, n_T)} \text{ and, therefore,}$$

$$u_{T-1}(\alpha_{T-1}, h_{T-1}) = \gamma_{T-1}^H \frac{1-(\varphi_T^L)^{h_{T-1}}}{1-\varphi_T^L} = \gamma_{T-1}^H \sum_{k=0}^{h_{T-1}-1} (\varphi_T^L)^k .$$

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

$$u_{T-1}(\alpha_{T-1},2) - u_{T-1}(\alpha_{T-1},1) = \gamma_{T-1}^H \varphi_T^L(\gamma_{T-1}^H,2) > 0 .$$

On the other hand,

$$\Delta_{T-1} \equiv \lim_{\alpha_{T-1}\to 1} \frac{u_{T-1}(\alpha_{T-1}, x) - u_{T-1}(\alpha_{T-1}, x+1)}{1-\alpha_{T-1}} = (1-q_{T-1}^L)(q_T^L(x) - q_T^L(x+1)) > 0 .$$

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 to take one low signal and to get zero pay-off. Thus, $l_{T-1}^* = 1 - \text{sign}(H_{T-1})$.

Ex-ante pay-off in period T-1 can now be written as

$$\begin{aligned}\overline{u_{T-1}} &= u_{T-1}(\alpha_{T-1},2)\cdot\left(1-(\varphi_{T-1}^L)^{n_{T-1}} - n_{T-1}(1-\varphi_{T-1}^L)(\varphi_{T-1}^L)^{n_{T-1}-1}\right) + \\ &\quad + u_{T-1}(\alpha_{T-1},1)\cdot n_{T-1}(1-\varphi_{T-1}^L)(\varphi_{T-1}^L)^{n_{T-1}-1} + \\ &\quad + u_{T-1}(\alpha_{T-1},0)\cdot(\varphi_{T-1}^L)^{n_{T-1}} \\ &= \alpha_{T-1}\left(\sum_{k=0}^{n_{T-1}-1}(\varphi_{T-1}^L)^k + \frac{(1-\alpha_{T-1})(1-q_{T-1}^L(n_{T-1}))}{1-\varphi_{T-1}^L}\left(\sum_{k=0}^{n_{T-1}-1}(\varphi_{T-1}^L)^k - n_{T-1}(\varphi_{T-1}^L)^{n_{T-1}-1}\right)q_T^L(2)\right).\end{aligned}$$

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

$$\overline{u_t}(\alpha_t, n_t) = \alpha_t \sum_{k=0}^{n_t-1}(\varphi_t^L)^k + \alpha_t \frac{(1-\alpha_t)(1-q_t^L(n_t))}{1-\varphi_t^L}\left(\sum_{k=0}^{n_t-1}(\varphi_t^L)^k - n_t(\varphi_t^L)^{n_t-1}\right)\beta_t ,$$

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 $l_t^* = 1 - \text{sign}(H_t)$ for all $\alpha_t \in (\alpha_t^*, 1)$. The corresponding reduced form pay-off function in stage t-1 is given by $u_{t-1}(\alpha_{t-1}, h_{t-1}) = \overline{u_t}(\alpha_t, h_{t-1})$. Suppose that $H_{t-1} \geq 2$. Then, it is easy to see that $u_{t-1}(\alpha_{t-1},2) > u_{t-1}(\alpha_{t-1},1)$:

$$u_{t-1}(\alpha_{t-1},2) - u_{t-1}(\alpha_{t-1},1) = \gamma_{t-1}^H(1-\gamma_{t-1}^H)(q_t^L(2) + (1-q_t^L(2))\beta_t) > 0 .$$

On the other hand,

$$\Delta_{t-1} = \lim_{\alpha_{t-1}\to 1} \frac{u_{t-1}(\alpha_{t-1},x)-u_{t-1}(\alpha_{t-1},x+1)}{1-\alpha_{t-1}} = \left(q_t^L(x)-q_t^L(x+1)\right)\left(1-q_{t-1}^L\right)\left(1-\beta_t\right) > 0 .$$

Thus, there exists 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 $l_{t-1}^* = 1-\text{sign}(H_{t-1})$ 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

$$\begin{aligned}
\overline{u_{t-1}}(\alpha_{t-1},n_{t-1}) &= u_{t-1}(\alpha_{t-1},2)\cdot\left(1-\left(\varphi_{t-1}^L\right)^{n_{t-1}} - n_{t-1}\left(1-\varphi_{t-1}^L\right)\left(\varphi_{t-1}^L\right)^{n_{t-1}-1}\right) + \\
&\quad + u_{t-1}(\alpha_{t-1},1)\cdot n_{t-1}\left(1-\varphi_{t-1}^L\right)\left(\varphi_{t-1}^L\right)^{n_{t-1}-1} + \\
&\quad + u_{t-1}(\alpha_{t-1},0)\cdot\left(\varphi_{t-1}^L\right)^{n_{t-1}} \\
&= \alpha_{t-1}\sum_{k=0}^{n_{t-1}-1}\left(\varphi_{t-1}^L\right)^k + \alpha_{t-1}\frac{\left(1-\alpha_{t-1}\right)\left(1-q_{t-1}^L(n_{t-1})\right)}{\left(1-\varphi_{t-1}^L\right)}\left(\sum_{k=0}^{n_{t-1}-1}\left(\varphi_{t-1}^L\right)^k - n_{t-1}\left(\varphi_{t-1}^L\right)^{n_{t-1}-1}\right)\beta_{t-1},
\end{aligned}$$

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 the induction, for any t=1,…,T-1 there exists an $\alpha_t^* \in (0,1)$ such that for all $\alpha_t \in \left(\alpha_t^*,1\right)$ and for all $\tau \geq t$: $h_\tau^* = \min(2,H_\tau)$ and $l_\tau^* = 1-\text{sign}(H_\tau)$. 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:

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

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 $h_1^* \leq \bar{n}$. ■