



**HAL**  
open science

# Everyday econometricians: Selection neglect and overoptimism when learning from others

Kai Barron, Steffen Huck, Philippe Jehiel

► **To cite this version:**

Kai Barron, Steffen Huck, Philippe Jehiel. Everyday econometricians: Selection neglect and overoptimism when learning from others. 2022. halshs-03735640

**HAL Id: halshs-03735640**

**<https://shs.hal.science/halshs-03735640>**

Preprint submitted on 21 Jul 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



WORKING PAPER N° 2022 – 20

Everyday econometricians: Selection neglect and overoptimism  
when learning from others

Kai Barron  
Steffen Huck  
Philippe Jehiel

JEL Codes: C11, C90, D80, D83

Keywords: Bounded rationality, selection neglect, beliefs, overconfidence,  
survivorship bias.



# Everyday econometricians: Selection neglect and overoptimism when learning from others\*

Kai Barron<sup>†</sup>, Steffen Huck<sup>‡</sup>, Philippe Jehiel<sup>§</sup>

January 24, 2019

## Abstract

There are many important decision problems where learning through experimentation is costly or impossible. In these situations, individuals may try to learn from observing the outcomes of others who have made similar decisions. Often, however, information about others comprises a selected dataset, as outcomes are observed conditional on specific choices having been made. In this paper, we design an investment game which allows us to study the influence of selection when learning from others. Using the theoretical study of selection neglect in Jehiel (2018) as a guide, we test (i) for the presence of selection neglect in this investment context, and (ii) some comparative static predictions of the model. We find strong evidence for selection neglect—even though subjects are fully informed about the data generating process. As theoretically predicted, the degree of bias due to selection neglect increases when other decision makers become more informed, or become more rational. It decreases when signals are correlated.

**JEL Codes:** C11, C90, D80, D83

**Keywords:** Bounded rationality, selection neglect, beliefs, overconfidence, survivorship bias.

---

\*The authors would like to thank Philipp Albert, Antonio Cabrales, Daniel Friedman, Thomas Graeber, Andrew Schotter, Georg Weizsäcker, Florian Zimmermann, the participants at several conferences, as well as numerous colleagues at the WZB for helpful comments and suggestions. Furthermore, we thank Markus Konrad for invaluable help with programming in o-Tree, and Nina Bonge for excellent support in running the experiments.

<sup>†</sup>WZB Berlin: kai.barron@wzb.eu

<sup>‡</sup>University College London and WZB Berlin: steffen.huck@wzb.eu.

<sup>§</sup>University College London and Paris School of Economics: jehiel@pse.ens.fr.

# 1 Introduction

Many important decisions are made infrequently over the course of a lifetime. These decisions do not permit much practise or experimentation and are costly to reverse. Such decisions include the choice of one's career, whether and where to buy a house, where to attend college, whether to switch jobs, whether to start a business, whether to get married, or whether to have children. A natural way to make these decisions is to consider the outcomes of others who have made the same decisions in the past. However, this comes with a fundamental challenge: data on the outcomes of a particular choice is only available for those who have made this choice.

The econometrics literature on selection has demonstrated that the average outcomes of individuals who have made a specific decision can differ substantially from the expected outcome of the average individual were she to make the same decision (see, for example, [Heckman, 1979, 1990](#)).<sup>1</sup> This is the case even after conditioning on a host of observed variables. Therefore, considering the outcomes of past generations (or members of one's peer group) who faced the same decision as a guide to making one's own decisions requires skillful interpretation of selected datasets. Neglecting this selection can lead to suboptimal decision making.

Specifically, we study experimentally the decision making of individuals who have to decide whether or not to invest in a project based on the observation of a private signal that can be thought of as the initial impression delivered by the project. A key feature of the experiment is that subjects have access to a selected database about past implemented projects that collects information about the success rate of past projects as well as about how the success rate varies with the signal. Importantly, the database is restricted to implemented projects as it seems plausible that in real life situations decision makers would not even be aware of the counterfactual outcomes of non-implemented projects. In our experiment, subjects are also informed of the data generating process, which in principle allows them to compute the probability of success of the project conditional on the signal with no need to rely on the database.

A question of interest is then whether subjects rely on the database and whether, if they do, they correct for selection when deciding on their investment strategy and, if they do not, whether the resulting decision making induces distortions away from the first-best. This question is all the more relevant in the investment context as there has been ample evidence that entrepreneurs exhibit an overoptimism bias regarding the chance of success of their business (see, e.g., [Cooper et al., 1988](#);

---

<sup>1</sup>The literature in econometrics generally discusses this problem from the perspective of the econometrician conducting statistical analysis. However, we are all 'everyday econometricians' when we use the information that we receive in our daily lives to draw inferences and guide our decision making. The objective of this paper is to conduct an exploration of the implications of selection neglect on the decision making of the average person.

Malmendier and Tate, 2005; Koellinger et al., 2007; Dawson et al., 2014). While such evidence is sometimes used to suggest that entrepreneurs may have a hard-wired form of overconfidence or that the bias is the result of a motivated cognition process, it is of interest to explore whether such biases could be the mere consequence of selection neglect.

Jehiel (2018) studies a theoretical model with this application in mind assuming that subjects rely on the database and extrapolate from it as if the sample were unbiased. We will refer to such subjects as selection neglect agents. Key findings in Jehiel (2018) are that, in the steady state, overinvestment that looks like it stems from an overoptimism bias results when the signals received by different subjects about the same projects are not perfectly correlated, and that the bias is exacerbated when decision makers are surrounded by investors who make better decisions. Thus, overoptimism arises as a consequence of selection neglect (unless signals are perfectly correlated), and the magnitude of the bias is affected by the composition of the pool of investors.

With these background considerations in mind, we design an investment experiment with various treatments that differ according to whether the signals received by different investors are correlated or not. In our treatment SELECTED (which is indicative of more complex investment decisions) signals are uncorrelated and in our treatment CORRELATED (which is indicative of simpler investment problems) signals are identical. Additionally, we implement the treatment EXTERNALITY where some computerized agents make smarter decisions. We also consider a control treatment, CONTROL, in which the database consists of all projects, including those that were not invested in.

Our main findings are as follows. First, we observe vastly different investment rates in the various treatments with the highest rate in EXTERNALITY, followed by SELECTED and, lastly, CORRELATED and CONTROL which exhibit investment rates close to those predicted by theory. This provides direct evidence that subjects make use of the database, and do not only rely on their information about the data generating process, as otherwise, the investment strategies would be the same across all treatments. Second, the comparative statics between treatments are in line with the theoretical predictions made in Jehiel (2018) in which subjects are assumed to extrapolate naively from the database as if it were unbiased.

Thus, even in conditions in which selection neglect could be bypassed (given the availability of the data generating process), subjects seem to naively extrapolate from observed data, and in some circumstances, such extrapolations give rise to an overoptimism/overinvestment bias, which turns out to be affected by the composition of the observed pool of projects in the database. For practical purposes, our results suggests that we should see less overoptimism in simple investment problems (in which signals are more likely to be correlated) than in complex investment problems

(in which signals are more likely to be idiosyncratic) and overoptimism should be exacerbated among junior investors when there are more senior investors around (viewing the latter as making better decisions). In future work, we plan to explore these predictions in the field.

This paper can be viewed as contributing to the fast growing literature studying biases in reasoning and belief formation in two ways. Firstly, it provides a new perspective on a bias that has received considerable attention in the literature, *overconfidence*.<sup>2</sup> As is typical in the literature considering biases, one can dichotomise the studies of overconfidence into two broad strands, (i) “wired-in” *cognitive mistakes* (see, e.g., [Malmendier and Tate \(2005\)](#); [Koellinger et al. \(2007\)](#); [Puri and Robinson \(2007\)](#); [Moore and Healy \(2008\)](#); [Logg et al. \(2018\)](#)), and, (ii) *motivated beliefs and reasoning* (see, e.g., [Compte and Postlewaite \(2004\)](#); [Von Hippel and Trivers \(2011\)](#); [Eil and Rao \(2011\)](#); [Möbius et al. \(2014\)](#); [Schwardmann and Van der Weele \(2016\)](#); [Zimmermann \(2018\)](#)). The current paper differs from both these strands in that it demonstrates how a distinct cognitive statistical bias, selection neglect, can generate behaviour that looks like overconfidence in certain situations. Importantly, the theory provides a clear description of the contextual factors that should enhance or reduce the presence of overconfidence. This is tested in the experiment.

Secondly, we contribute to the literature studying cognitive statistical biases, enriching the understanding of how people draw inference and form beliefs when observing data with a specific structure. Prior research in this area includes work by [Enke and Zimmermann \(2017\)](#), who study how individuals form beliefs when they observe correlated signals, which can result in mistaken “double-counting” of the same information, or *correlation neglect*. [Jin et al. \(2018\)](#) study whether individuals accurately interpret the “absence” of a signal.<sup>3</sup> Similarly, [Graeber \(2018\)](#) studies the mistakes in belief formation that can result from individuals failing to account for the noise in the information structure.

The two experimental papers closest to ours, [Enke \(2017\)](#) and [Esponda and Vespa \(2018\)](#), study how individuals draw inferences when they observe data that is only observed conditional on satisfying a specific criterion. In line with the first set of results of the current paper, both find

---

<sup>2</sup>For example, Daniel Kahneman stated that if he could eliminate one human bias, he would get rid of overconfidence ([Malmendier and Taylor, 2015](#)); Werner DeBondt and Richard Thaler assert that “Perhaps the most robust finding in the psychology of judgement is that people are overconfident.” ([De Bondt and Thaler, 1995](#)); and, Roland Benabou and Jean Tirole write that “At an individual level, overconfidence is perhaps the most common manifestation of the motivated-beliefs phenomenon” ([Bénabou and Tirole, 2016](#)).

<sup>3</sup>Since observing no signal is technically contradictory, for clarification, we mean this in the colloquial sense of Sherlock Holmes’ dog that did not bark. [Jin et al. \(2018\)](#) study sellers who withhold information, and find buyers who underweight the information content of non-disclosure.

evidence that people fail to condition on the selection rule.<sup>4</sup> Experimental subjects interpret the data they observe as reflecting the unconditional distribution, thereby displaying *selection neglect*.<sup>5</sup> However, [Enke \(2017\)](#) and [Esponda and Vespa \(2018\)](#) consider very different environments to the current paper, with [Enke \(2017\)](#) focusing on “echo chambers” and [Esponda and Vespa \(2018\)](#) considering a voting environment where hypothetical thinking is important. Furthermore, neither studies how the implications of selection neglect can be exacerbated or ameliorated under different information structures, as in the current paper.

Finally, from a theory viewpoint, the theoretical approach developed in [Jehiel \(2018\)](#) and discussed later in the theoretical framework section can be related to the recent strand of literature on coarse reasoning first studied in [Jehiel \(2005\)](#), [Eyster and Rabin \(2005\)](#) and pursued in [Jehiel and Koessler \(2008\)](#) and [Esponda \(2008\)](#) among others (see [Spiegler’s \(2011\)](#) book). However, static approaches to coarse reasoning such as in [Eyster and Rabin \(2005\)](#) or [Jehiel and Koessler \(2008\)](#) are not sufficient to model steady states in which strategies are based only on data about implemented projects. Either the use of self-confirming equilibrium ideas as in [Esponda \(2008\)](#) or an extensive-form game approach to coarse reasoning as in [Jehiel \(2005\)](#) facilitates the formalization of the idea that subjects would extrapolate from success rate data about implemented projects to all projects irrespective of whether or not they were implemented (see the CEPR WP of [Jehiel \(2018\)](#) for details).

The remainder of the paper is organised as follows. Section 2 provides the theoretical framework. Section 3 outlines the experimental design. Section 4 presents the results. Section 5 contains a discussion, and Section 6 concludes.

## 2 Theoretical Framework

Similarly to [Jehiel’s \(2018\)](#) model, we consider the investment problems of a population of risk neutral agents where each investor  $i$  faces a new project and must decide whether to *invest* or *not*

---

<sup>4</sup>In a noteworthy earlier contribution to the topic, [Koehler and Mercer \(2009\)](#) study inference about mutual funds when companies selectively advertise only their best performing funds. The authors provide evidence that both experts and novice investors fail to take into account this selective advertising, using the term *selection neglect* to describe the phenomenon.

<sup>5</sup>It is perhaps important to note that *selection neglect* is similar in spirit to another cognitive statistical bias, namely the *law of small numbers* (see, e.g., [Tversky and Kahneman, 1971](#) and [Rabin, 2002](#)). Essentially, while *selection neglect* involves misinterpreting a conditional distribution as resembling an underlying unconditional distribution, the *law of small numbers* involves overestimating the degree to which a small but randomly selected sample resembles the true underlying distribution from which it is drawn. One might group *correlation neglect*, *selection neglect*, the *law of small numbers*, and *underweighting null signals* collectively as biases where individuals misinterpret the relationship between the signals they receive and the underlying data generating process.

*invest*. The cost of investing in a project is  $c > 0$ . The return of a project,  $x$ , is a binary random variable. Either the project is successful,  $x = \bar{x} > c$ , or the project is unsuccessful,  $x = \underline{x} = 0 < c$ .

When an investor faces a new project, she knows the cost of the project, but she doesn't know whether the project will be successful or unsuccessful. However, prior to making the investment decision, she receives a private, informative signal, about the project. One can think of this as the investor's private, informative *impression* about the project's *ex ante* likely success.<sup>6</sup> (Note, we will use the terms "signal" and "impression" interchangeably.) Given this signal, the investor must decide whether to invest or not.

To account for the possibility that different agents may look at different characteristics of projects to form their impression, we model the statistical link between signals and success as follows. (It will fit the experimental setting to be described next.)

The success  $x$  of a project is fully determined by a triple  $(s_A, s_B, s_C)$  where  $s_A, s_B$  and  $s_C$  are the realizations of independent discrete uniform distributions on  $\{1, \dots, 10\}$ . Think of each  $s_X$  as the realization of a fair ten-sided dice. The agent facing a new project will see the realization of one of  $s_A, s_B$  or  $s_C$  but not the realizations of the other two signals. So there are three possible types of agents: The  $A$  agents observing the  $s_A$  signal; the  $B$  agents observing  $s_B$  signal, and the  $C$  agents observing the  $s_C$  signal.

In all cases, we assume that whether a project is successful or not is determined by  $(s_A, s_B, s_C)$  according to the rule

$$x = \begin{cases} 0 & \text{if } W < 22 \\ \bar{x} & \text{if } W \geq 22 \end{cases}$$

where  $W = s_A + s_B + s_C$ . That is, the project is successful if the total score over the three dimensions as measured by  $W$  is above the threshold 22.

The central question of interest here is how the investor decides on her investment strategy, and in particular how she forms a belief about the relationship between the signal she observes and the likely success of the project she faces. We study investor behaviour under four different information regimes, comparing the behaviour of our experimental participants to two benchmark agents: (i) the rational Bayesian decision maker, and (ii) the Selection Neglect investor as considered in [Jehiel \(2018\)](#) who extrapolates from the observed database as if it were unbiased.

---

<sup>6</sup>We use the term impression as it may reflect the investor's (informative) gut feeling about a project.



## The feedback environment

In our main treatments, subjects are informed of the statistical model linking the signals to the return. However, in addition to this, we also provide subjects with information about past implemented projects. Importantly, subjects only receive information about projects that were invested in, not about the other projects. Moreover, for all those projects, an agent of type  $A$  (resp.  $B$ , or  $C$ ) gets access to the realization of the signal  $s_A$  (resp.  $s_B$  or  $s_C$ ) of previously implemented projects as well as the corresponding return. An agent is not informed of the signal that was received by the agent who decided to invest in the project.<sup>7</sup>

Notice that feedback is irrelevant for a **rational Bayesian decision maker** who can make the best decision simply based on the statistical model, but it will affect the behaviour of Selection Neglect investors who assess the link between their impression and the return by extrapolating from the data seen in their feedback. As it turns out, for those agents referred to as **Selection Neglect investors**, the composition of the types in the population of investors affects the form of their behaviour. We will consider four scenarios.

### *Scenario 1: Learning from others*

Our first main scenario is one in which a type  $A$  agent only observes projects that were invested in by type  $B$  or  $C$  agents. This is meant to represent situations in which investment decisions are complex and the chance that two different investors would consider the same dimension of the project (to form their impression) is small.

### *Scenario 2: Learning from others with correlated signals*

Our second scenario will consider the case in which all investors are of the same type, either representing situations in which only one dimension can be scrutinized prior to the investment decision or situations in which investors are homogeneous for whatever reason in their way of forming a first opinion about projects. There are many possible reasons why impression formation may be more homogeneous in some contexts than in others. For example, one might expect impression formation to be more homogeneous in (i) simple contexts rather than complex contexts, (ii) when the individuals have a similar world view, rather than having vastly differing perspectives, or (iii) when one particular dimension of the project is made salient to all individuals.

---

<sup>7</sup>Think of this, for example, as an investor always looking at a project through the lens of her subjective personal impression. She may form an impression or receive a signal that is equally informative about the likely success of the project as the signal received by the individual who made the investment choices, but since the two individuals may focus on different dimensions of the situation, their signals may differ. This is particularly likely to be the case when the situation is complex. In simple contexts, impression formation is more likely to be done similarly by different individuals.

### *Scenario 3: Learning from better informed individuals*

A third scenario that we consider has investors of a given type  $A$  facing in their feedback either projects that were handled by investors of another type  $B$  or  $C$  or by omniscient investors that would invest only if the project is successful (the latter are implemented through machines in the experiment). This scenario is meant to represent situations in which less experienced investors would be exposed in their feedback to a mix of more or less experienced investors.

### *Scenario 4: Learning from others when also observing the counterfactual*

As a benchmark for comparison, we consider a scenario in which agents receive feedback on all projects — those that were invested in, as well as those that weren't invested in. In this scenario, a type  $A$  agent will receive feedback on all projects faced by type  $B$  or  $C$  agents. The choices made by these other agents are irrelevant for the feedback she receives, and consequently there is no selection of feedback. Aside from serving as a comparative benchmark, this scenario captures situations in which an individual makes a decision, but the outcome will be observed irrespective of the decision.<sup>8</sup>

## **Bayesian benchmark**

Given the symmetry of the three types of signals, let us consider  $s_A$  as a representative signal. A Bayesian can compute  $P(\bar{x}|s)$ , the Bayesian posterior probability of success conditional on  $s_A = s$ . The above statistical model gives rise to  $P(\bar{x}|s) = \frac{1}{200} \cdot (s^2 - s)$ . Thus, given the assumed risk neutrality, a Bayesian investor follows the decision rule of investing whenever she observes a signal  $s$  satisfying  $s \geq s^{Bayes}$  where  $s^{Bayes}$  is the smallest  $s$  satisfying

$$P(\bar{x}|s) > \frac{c}{\bar{x}} \iff \frac{1}{200} \cdot (s^2 - s) > \frac{c}{\bar{x}}. \quad (1)$$

In all scenarios, we have that  $s^{Bayes} = 9$  when  $\bar{x} = 3.40$  and  $c = 1$  (which are the parameters we use in the experiment).<sup>9</sup>

---

<sup>8</sup>This would, for example, include any decision that involves a bet on some observable event. For example, buying a house, which involves a bet on the performance of the housing market. One would later be able to observe whether it would have been a good decision, irrespective of the decision made.

<sup>9</sup>The Bayesian investor doesn't rely on her feedback to formulate her investment strategy, and therefore follows this threshold strategy in all four scenarios.

## Equilibrium with selection neglect investors

As an alternative to the Bayesian benchmark, we consider the case in which an investor of type  $A$  would consider the proportion  $\hat{p}$  of successful projects in her feedback with the same realization  $s_A$  as her current project and invest if  $\hat{p} \cdot \bar{x} > c$  and not invest otherwise. We refer to such investors as Selection Neglect investors as they do not internalize that the information they see is biased due to the fact that they only see projects that were invested in.

In [Jehiel \(2018\)](#), an equilibrium is proposed to describe the steady state generated by such a heuristic investment strategy. To present the equilibrium, we consider the idealized situation with a continuum of investors and we let  $q_{inv}(s)$  with  $s = (s_A, s_B, s_C)$  denote the probability with which a project with characteristics  $(s_A, s_B, s_C)$  was invested in in the past in the pool observed by an investor of type  $A$ .

A Selection Neglect investor of type  $A$  observing signal  $s_A = s^*$  would think that the probability of success of such projects is

$$\hat{P}(\bar{x}|s^*; q_{inv}) = \frac{\sum_{s_B, s_C} 1_{s^* + s_B + s_C \geq 22} \cdot q_{inv}(s^*, s_B, s_C)}{\sum_{s_B, s_C} q_{inv}(s^*, s_B, s_C)} \quad (2)$$

where  $1_{s^* + s_B + s_C \geq 22} = 1$  if  $s^* + s_B + s_C \geq 22$  and 0 otherwise given that all  $(s_B, s_C)$  are equally likely and only those projected invested in are observed. Accordingly, a Selection Neglect investor will invest if  $\hat{P}(\bar{x}|s^*; q_{inv}) \cdot \bar{x} > c$  and not invest otherwise.

In equilibrium, the investment strategy of the various types of investors should give rise to  $q_{inv}$  which leads to a fixed point formulation that depends on which scenario we are in. The following describes the equilibrium in each scenario, with further details and intuition provided in [Appendix A](#).

### *Scenario 1: Learning from others*

In this case, a symmetric pure strategy equilibrium would require that an investor of type  $X$  invests if she gets a signal  $s_X \geq s^{SN1}$  and thus  $q_{inv}^{SN1}(s^*, s_B, s_C) = 1$  if  $s_B, s_C \geq s^{SN1}$ ,  $q_{inv}^{SN1}(s^*, s_B, s_C) = 0.5$  if  $s_B \geq s^{SN1} > s_C$  or  $s_C \geq s^{SN1} > s_B$  and  $q_{inv}^{SN1}(s^*, s_B, s_C) = 0$  otherwise where  $s^{SN1}$  should be such that  $\hat{P}(\bar{x}|s^{SN1}; q_{inv}^{SN1}) \cdot \bar{x} > c$  and  $\hat{P}(\bar{x}|s^{SN1} - 1; q_{inv}^{SN1}) \cdot \bar{x} < c$  if  $s^{SN1} > 1$ .

Given the symmetry of the problem, it is readily verified that  $\widehat{P}(\bar{x}|s_A; q_{inv}^{SN1})$  simplifies into  $P(\bar{x}|s_A; s_B \geq s^{SN1})$  which is clearly larger than  $P(\bar{x}|s_A)$  (because the extra conditioning on  $s_B \geq s^{SN1}$  shifts upwards the probability of success). This in turn implies that there is more investment in the equilibrium with coarse investors in scenario 1. We have  $s^{SN1} = 6$  when  $\bar{x} = 3.40$  and  $c = 1$ .

### *Scenario 2: Learning from others with correlated signals*

When all investors are of type  $A$ , if agents follow the threshold strategy to invest if  $s_A$  is no smaller than  $s^{SN2}$ , we would have  $q_{inv}^{SN2}(s^*, s_B, s_C) = 1$  if  $s^* \geq s^{SN2}$  and 0 otherwise. This implies that for all  $s_A \geq s^{SN2}$ ,  $\widehat{P}(\bar{x}|s_A; q_{inv}^{SN2}) = P(\bar{x}|s_A)$  and thus it cannot be that  $s^{SN2} < s^{Bayes}$  (given that  $P(\bar{x}|s_A)\bar{x} < c$  for any  $s_A < s^{Bayes}$ ). When signals are perfectly correlated among investors, there cannot be overinvestment in equilibrium, and imposing some exogenous trembling would force  $s^{SN2} = s^{Bayes}$ .

### *Scenario 3: Learning from better informed individuals*

In this case investors of a type  $A$  face in their feedback projects that were handled either, by coarse investors of type  $B$  or  $C$ , or, by omniscient investors that would invest only if the project is successful (the latter are implemented through machines). Letting  $\lambda$  denote the proportion of non-omniscient investors, and letting  $s^{SN3}$  denote their equilibrium threshold, one would have  $q_{inv}^{SN3}(s^*, s_B, s_C) = \lambda q^C(s^*, s_B, s_C) + (1 - \lambda)q^R(s^*, s_B, s_C)$  where  $q^C(s^*, s_B, s_C) = 1$  if  $s_B, s_C \geq s^{SN3}$ ,  $q^C(s^*, s_B, s_C) = 0.5$  if  $s_B \geq s^{SN3} > s_C$  or  $s_C \geq s^{SN3} > s_B$  and  $q^C(s^*, s_B, s_C) = 0$  otherwise,  $q^R(s^*, s_B, s_C) = 1$  if  $s^* + s_B + s_C \geq W$  and 0 otherwise, and  $s^{SN3}$  should be such that  $\widehat{P}(\bar{x}|s^{SN3}; q_{inv}^{SN3}) \cdot \bar{x} > c$  and  $\widehat{P}(\bar{x}|s^{SN3} - 1; q_{inv}^{SN3}) \cdot \bar{x} < c$  if  $s^{SN3} > 1$ .

It can be shown that when there are more fully informed omniscient investors around (when  $\lambda$  is smaller), the overinvestment bias increases, i.e.,  $s^{SN3}$  gets smaller. When  $\lambda = 1/2$ ,  $\bar{x} = 3.40$  and  $c = 1$  (as in our experiment), we have a symmetric equilibrium in mixed strategies, where players mix between playing  $s^{SN3} = 5$  with probability  $\mu = 0.8$  and playing  $s^{SN3} = 6$  with probability  $1 - \mu = 0.2$ .<sup>10</sup>

---

<sup>10</sup>Note,  $\lambda$  refers to the fraction of fully informed omniscient investors amongst those who are generating the feedback. This is why  $\lambda = 1/2$  is relevant in relation to our experimental design, and not  $\lambda = 1/3$ .

#### *Scenario 4: Learning from others when also observing the counterfactual*

Here, type  $A$  investors observe feedback about all projects faced by type  $B$  or  $C$  investors, irrespective of whether they invested or not. This implies that while Selection Neglect investors still form their belief about the mapping from signals to success probabilities,  $\widehat{P}(\cdot|\cdot)$ , according to equation 2, they now observe the outcomes of past projects with probability one. Therefore,  $q_{inv}$  should be replaced by  $q = 1$  in equation 2, and Selection Neglect investors form a belief  $\widehat{P}(\bar{x}|s^*; q = 1)$ , which in expectation is equal to the Bayesian posterior,  $P(\bar{x}|s^*)$ . Consequently,  $s^{SN4} = s^{Bayes}$ .

### **Summary of hypotheses**

The theoretical predictions for our two benchmark agents, the rational Bayesian and the Selection Neglect investor, in each of the four scenarios we consider is summarised in Table 2. The table provides the threshold strategies in each scenario, as well as the propensity to invest.<sup>11</sup> The latter provides a convenient outcome measure and is used in the main empirical analysis below.

Our experiment is designed to reflect these four scenarios, with each treatment representing one scenario. Table 2 provides clear guidance for our hypotheses in the experiment, since the Bayesian agent has the same propensity to invest in all four treatments, while the coarse investor's propensity to invest can be ranked across the four scenarios. We have three main hypotheses, each of which tests one consequence of selection neglect against the alternative hypothesis that the average subject's behaviour is approximately Bayesian:

**Hypothesis 1.** *Due to the influence of being exposed to selected data, participants will invest more in Scenario 1 than in Scenario 4.*

**Hypothesis 2.** *Due to the negative (information) externality exerted by highly informed decision makers, the propensity to invest in Scenario 3 will be higher than in Scenario 1.<sup>12</sup>*

**Hypothesis 3.** *The increased correlation in the signals in Scenario 2 implies that the influence of selection neglect will be ameliorated, and the propensity to invest will be lower than in Scenario 1, and roughly equal to Scenario 4, even though subjects in Scenario 2 only observe projects that are invested in.*

---

<sup>11</sup>Since each of the ten signals is equally probable, the investment propensity follows directly from the threshold strategy.

<sup>12</sup>Note, in order to focus directly on the influence of the externality exerted by exogenously increasing the investment quality of a subset of individuals, this hypothesis is referring to the average investment propensity, when *excluding the omniscient investors*. This allows for a direct comparison with the investors in Scenario 1, as any change in investment must operate through the investment outcomes they are observing.

Table 1: Summary of theoretical predictions across scenarios

	Threshold	Investment Propensity
<u>Selection Neglect agent</u>		
<i>Scen. 1: SELECTED</i>	$s^{SN1} = 6$	0.5
<i>Scen. 2: CORRELATED</i>	$s^{SN2} = 9$	0.2
<i>Scen. 3: EXTERNALITY</i>	$s^{SN3} = 5$ with $\mu = 0.8$ $s^{SN3} = 6$ with $1 - \mu = 0.2$	0.58
<i>Scen. 4: CONTROL</i>	$s^{SN4} = 9$	0.2
<u>Bayesian agent</u>		
<i>Scen. 1 - 4: ALL TREATMENTS</i>	$s^{Bayes} = 9$	0.2

### 3 Experimental Design and Procedures

The objective of the experimental design is to construct an environment in which the influence of observing only projects that are invested in can be tested while varying the degree of sophistication of the decisions makers and/or the correlation of signals observed by different subjects about the same project.

This has led us to consider the environment described above in which whether a project’s success is determined by the realization of three dice, and private information is modelled as the observation of one dice. We provide now details of the experimental design, echoing the main features of the model presented in Section 2.<sup>13</sup>

#### A Single Investment Decision

When an investor faces a new project, she makes a binary choice: to *invest* or *not invest*. She alone has the opportunity to invest in that particular project — i.e. it is not a group activity. The success of a project is determined by rolling three fair ten-sided dice, with the sides taking values 1 to 10.

<sup>13</sup>The instructions for the SELECTED treatment are included in the Appendices.

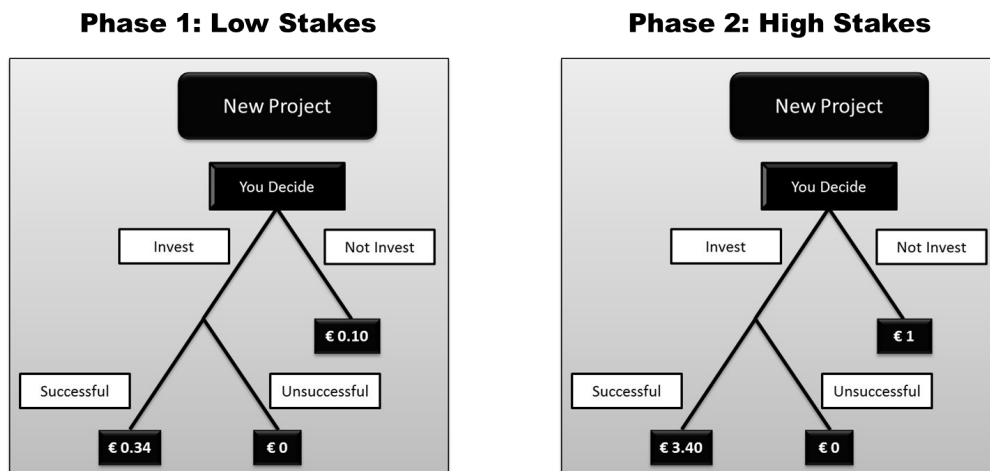
A project is *successful* ( $x = \bar{x}$ ) when the sum of these three dice is weakly greater than 22, and *unsuccessful* ( $x = 0$ ) otherwise.

Before making the investment decision, the participant observes the value of one of the three dice a, b, c.<sup>14</sup> The observed dice value is informative about the likely success of the project, with higher dice values implying higher success probabilities. A dice a, b or c is referred at as an attribute in the experiment.

### Costs and Benefits of an Investment

Every subject faces 20 rounds of investment decisions. In the first 10 rounds (Low Stakes), choosing to invest has a cost of €0.10, a *successful* investment pays out €0.34, while an *unsuccessful* investment pays out €0. In rounds 11 to 20 (High Stakes), choosing to invest has a cost of €1, a *successful* investment pays out €3.40, while an *unsuccessful* investment pays out €0. In each round, participants are endowed with money to cover the cost of making an investment in that round. This ensures that they cannot earn a negative payoff. The possible outcomes of an investment decision are summarised in Figure 1 below (which was also displayed to participants in the instructions for the experiment). The rationale for the Low Stakes phase is that it provides a period during which participants can learn about how the investment game works, while also building a large database of past investment outcomes to be used for feedback. Note that with risk-neutral agents, the two specifications (High stakes vs Low stakes) would make no difference to investment choices.

Figure 1: Overview of the possible outcomes of an investment decision



<sup>14</sup>This dice value is described to participants as being an *attribute* of the project. This *attribute* assumes the role of the *private impression*,  $s_i$ , described in the theoretical section above.

### *Number of Investment Decisions*

An important objective of the experimental design is allowing subjects to learn from the decisions of others, while avoiding confounding issues that may arise from learning from small samples. However, there is a tension between generating a large database of investment decisions and avoiding participant fatigue from making too many repetitive decisions. We resolve this issue using the strategy method (Selten, 1967) in the following way.

In each of the 20 rounds, a subject must decide for each of the ten possible attribute values whether she would invest or not, defining her *investment plan* for that round. The computer then randomly generates 50 projects (i.e. 50 sets of three dice rolls) for that specific subject in that round. Using her *investment plan*, each of the 50 projects is then assessed and investments made where prescribed by the strategy. In each round, one project is randomly chosen to be relevant for payment.

### **What information can subjects use to guide their investment decisions?**

A key object of interest in the investment decision is the participant’s belief about the probability that a project will be successful, conditional on each possible attribute value (i.e. her belief about  $P(\bar{x}|s_A)$ ). Participants in the experiment have access to two sources of information to guide their assessment of this probability:

1. A full description of the data generating process (**DGP**),
2. Information from past investments made by others (“**Learning from others**”).

### *Using information about the data generating process (DGP)*

In the experiment, participants were provided with a detailed description of the DGP.<sup>15</sup> For the Bayesian benchmark agent, this information about the DGP is sufficient to calculate  $P(\bar{x}|s_A)$ , and hence the optimal investment strategy. Therefore, information received about others’ outcomes is completely superfluous once she is informed about the DGP, so she will ignore the information received from others.

---

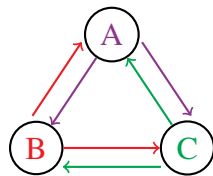
<sup>15</sup>The experimental instructions carefully described how a project is comprised of three fair ten-sided dice rolls (Dice **a**, Dice **b**, and Dice **c**) and that if the sum was weakly greater than 22, the project would be *successful*.



## *Learning from Others*

In contrast to the Bayesian agent, the Selection Neglect investor relies on the data she receives about the outcomes of investments made by others. This information is generated in the following way.

Participants are randomly assigned to groups of 3 participants. They stay in this group of 3 throughout all twenty rounds of the investment game. Each participant in the group is given a player label, namely Group Member **A**, Group Member **B**, Group Member **C**. Within each group, each player is able to observe the outcomes of past investments made by the two other group members (but not the outcomes of their own investments).



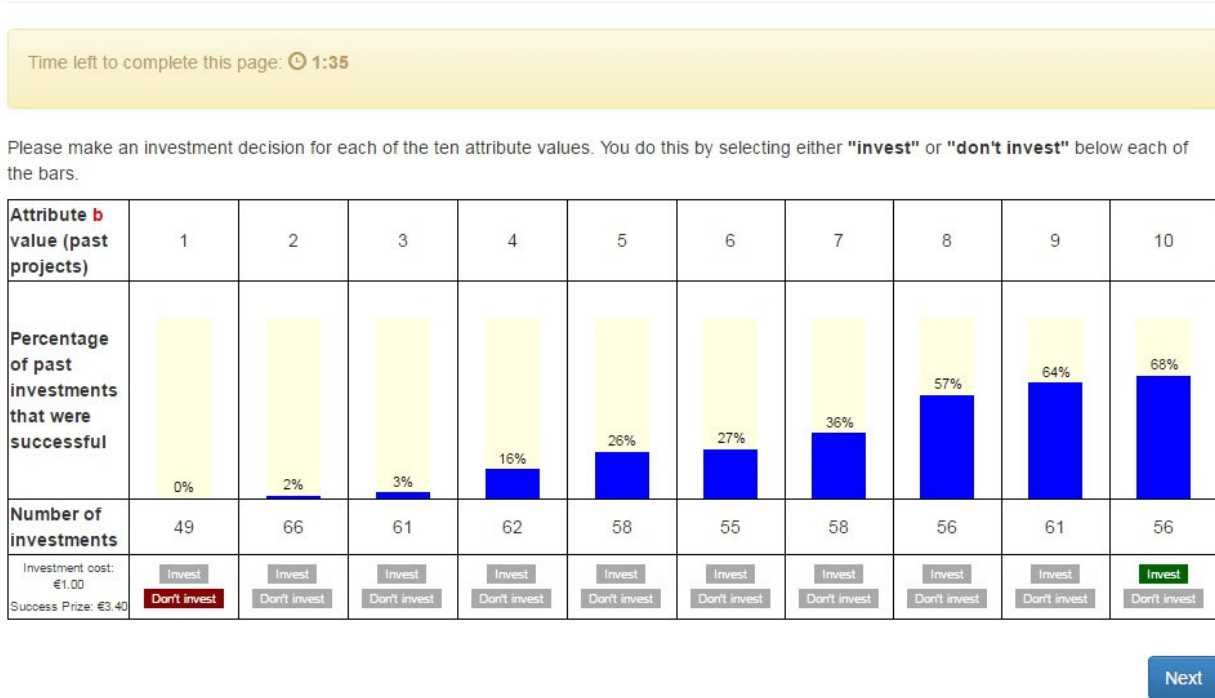
## *Updating the Personal Database*

Over the course of the experiment, participants receive a large quantity of data regarding the outcomes of past projects (up to 2000 projects in total).<sup>16</sup> In order to assist them, the information each subject receives is organised into an infographic containing the information in their *personal database*. Figure 2 provides an example of an infographic that a Group Member **B** might see. All projects the individual receives are collected into buckets according to the relevant attribute (in this case, attribute **b**). For each attribute value from 1 to 10, the infographic reports the number of projects in each bucket, and the fraction of those projects that were successful.

<sup>16</sup>At the end of every round, each participant's database of observed past projects was updated by adding all projects that: (i) were part of the fifty projects faced by each of the other group members (one hundred in total); and (ii) were invested in by that group member. In the first round, all group members start with an empty database, and in every subsequent round, all projects that satisfy these two criteria are added to the participant's personal database. This implies that the number of observations in the database grows over time.

Figure 2: Screenshot of the data observed by a participant (an example)

## Round 15/20 (high stakes) - Please define your Investment Plan



Participants provide their *investment plans* in each round by clicking one of the two options, “invest” or “don’t invest”, at the bottom of every attribute value column.

### Treatments: the feedback scenarios

We consider four treatment conditions that differ only in the structure of the feedback. Each treatment replicates one of the four information scenarios described in the theory section above. The SELECTED treatment is our central treatment, and in each of the other three treatments one aspect of the structure of the feedback is altered slightly. However, in all four treatments, participants are provided with a full description of the DGP.

#### Scenario 1: SELECTED TREATMENT

The SELECTED treatment allows us to examine the investment behaviour of participants in feedback scenario 1, where they are provided with access to the outcomes of all projects invested in by their group members in earlier rounds. For example, type A group members observe the

investments of the type **B** and **C** individuals in their group. However, individuals always observe the attribute corresponding to their type (e.g., attribute **a** for a type **A** group member). Comparison with the theoretical benchmarks allows a classification of individuals as Bayesian or Selection Neglect investors. Furthermore, a comparison of investment behaviour under the different feedback scenarios allows us to detect the presence of selection neglect agents and test some comparative statics of the theory.

### *Scenario 2: CORRELATED TREATMENT*

The **CORRELATED** treatment is identical to the **SELECTED** treatment, with the exception that all three group members are of the same type and therefore observe perfectly correlated project attributes. This implies that when a participant receives feedback from her two group members, she now observes the same project attribute as her group member.

### *Scenario 3: EXTERNALITY TREATMENT*

The **EXTERNALITY** treatment provides a test of the theoretical prediction that Selection Neglect investors can be negatively affected by the presence of more efficient investors. We experimentally implement such a variation by replacing one member of every group by a computer player who has perfect information about all three dice rolls and therefore only invests in successful projects. Other than this, the **EXTERNALITY** treatment is identical to the **SELECTED** treatment. The decision rule that the computer player follows is simply: *invest if and only if the project will be successful*.<sup>17</sup> This rule is known to participants.<sup>18</sup> With one of the three players only investing in successful projects, the selection effect of feedback becomes stronger, implying that Selection Neglect investors should overinvest even more than in the **SELECTED** treatment.

---

<sup>17</sup>It is worth mentioning the rationale for this choice of decision rule as opposed to the natural alternative—i.e. having the computer behave as a fully Bayesian agent would, namely investing for an attribute value of 9 and 10, and not otherwise. There are several reasons for our choice. Firstly, in describing the alternative rule, we would need to tell subjects the optimal decision rule without providing an explanation for why the computer follows this rule. This would make following a threshold decision rule salient, as well as anchoring participants to the specific Bayesian threshold rule. Therefore, if they followed this rule, we would not know if they had recognised the effect of selection, or were simply anchored to the computer’s decision rule. Secondly, the computer decision rule that we use has the attractive features that it: (i) is easy to understand, (ii) has a more pronounced effect on the selection of projects, and (iii) makes it salient to participants that the computer is acting in an optimal way and selecting only successful projects. This should help participants to recognise selection and would work against finding an effect of the treatment.

<sup>18</sup>In particular, participants are told: “Unlike the two human players who observe only one attribute value, the Robot Player observes all three attribute values. Therefore, the Robot Player always knows when a project it is considering will be successful in advance. The Robot Player therefore always invests if the project will be successful, and never invests in unsuccessful projects.”

#### *Scenario 4: CONTROL TREATMENT*

To accurately measure the effect of being exposed to selected feedback, it is important to observe investment behaviour when the feedback is not selected. Our CONTROL treatment provides participants with feedback from *all* projects faced by their group members, irrespective of whether they invested or not. It is otherwise identical to the SELECTED treatment.

#### **Implementation of the Experiment**

With the aim of ensuring that subjects had a clear understanding of the decision problem they faced, at the beginning of the experiment, participants received detailed instructions. Once all participants had carefully read through the instructions, they were required to complete a comprehensive set of control questions in order to ensure that all subjects had fully absorbed and understood the instructions. Only after correctly answering all the control questions did participants move on to the investment game.<sup>19</sup> Additionally, the pattern of decisions described in the results section below suggests that subjects understood the decision problem that they faced and the individual level data indicates that the vast majority of the subjects followed a threshold strategy in each of the last 5 rounds. Furthermore, very few subjects invested for low attribute values.

The experiments were conducted at the WZB-TU laboratory in Berlin between December 2016 and January 2018, with two sessions of 24 subjects for each Treatment group. This implies a total of 8 sessions, with 192 participants (48 in each Treatment Group). Participants were solicited through an online database using ORSEE (Greiner, 2015) and the experiment was run using the experimental software, o-Tree (Chen et al., 2016).

After the 20 rounds of the investment game, there was also a risk elicitation task and we obtained some demographic and other non-incentivised measures.<sup>20</sup> Sessions lasted up to 90 minutes. Average earnings from the investment game were €17.8 (including a €5 show-up fee).

---

<sup>19</sup>For example, question 2 of the control questions aimed to ensure that subjects were aware that there was no role for experimentation: “Do you observe the outcome of projects that *you* invested in before the end of the experiment?”

<sup>20</sup>Additionally, after both, the investment game and the risk elicitation task, participants completed a single round public goods game that was completely unrelated to this study. The instructions for the public goods game were only handed out after the investment game was completed. This implies that while completing the investment game subjects only knew that they would play a second game that was completely unrelated to the investment game. This information was identical across treatments groups.

## 4 Results

In the theoretical framework section above, we discussed the behaviour of two benchmark agents: (i) a fully rational Bayesian agent, and (ii) a selection neglect agent. The predicted behaviour of these two benchmark agents differs starkly under the four feedback scenarios discussed, generating comparative static predictions for the two agents. Our experiment is designed such that each treatment mirrors one of these four information scenarios. This allows us to directly evaluate the comparative static predictions of the theory by comparing investment behaviour between treatments.

Recall that the Bayesian agent makes the same investment choices in all four information scenarios (i.e.  $\text{CORRELATED} = \text{CONTROL} = \text{SELECTED} = \text{EXTERNALITY}$ ), while the propensity to invest of the selection neglect agent can be ranked by treatment as follows:  $\text{CORRELATED} = \text{CONTROL} < \text{SELECTED} < \text{EXTERNALITY}$ . The overarching question we wish to study is whether the observed pattern of behaviour more closely reflects the predicted behaviour of the rational Bayesian or the selection neglect agent.

Figure 3: Investment fraction in last five rounds, by treatment

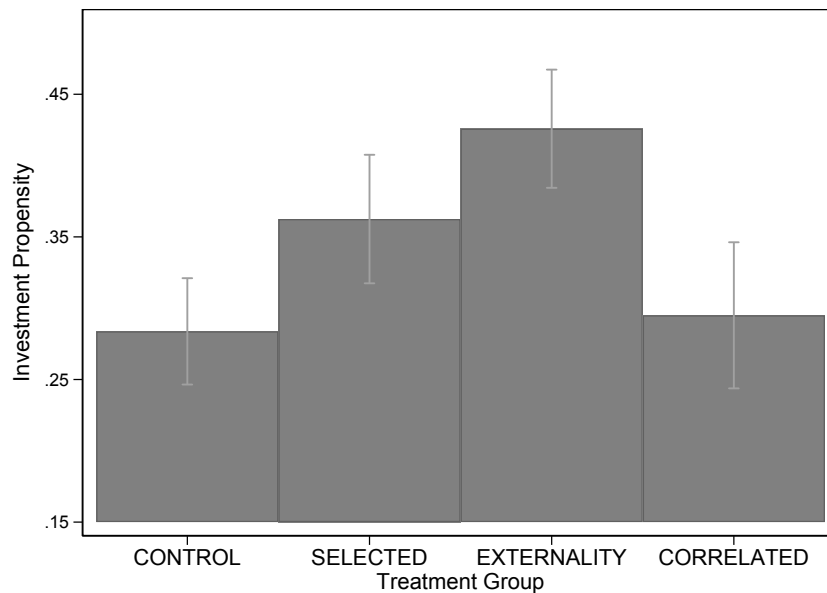


Figure 3 provides a first look at investment behaviour between treatments. Counting each individual as a single observation, the figure reports the average propensity to invest in the last

five rounds.<sup>21</sup> It is striking that the pattern of investment behaviour between treatment conditions strongly reflects the pattern predicted by the selection neglect agent model, with investment propensity between treatments ranked as follows: CORRELATED  $\approx$  CONTROL < SELECTED < EXTERNALITY. This evidence suggests that selection neglect is present amongst the participants in our experiment, leading to overinvestment. Below, we provide further evidence in support of this assertion.

### **Do we observe evidence of selection neglect?**

Here, we test the fundamental assumption underlying the theoretical framework, namely that there are individuals who fail to fully account for the way in which the data they observe has been selected. We test for the presence of selection neglect by comparing investment behaviour when participants are exposed to a selected database of past projects in SELECTED with behaviour in the CONTROL treatment. This is a test of Hypothesis 1.

Since we used the strategy method to elicit subjects' desire to invest for each attribute value in each round, we can directly obtain each participant's investment propensity by counting the number of attributes values for which she wished to invest within a given round. For example, if a subject chose to invest for all attribute values between 6 and 10 in a round, then her propensity to invest was 0.5. If instead she behaved as the Bayesian agent would and invested only for attributes 9 and 10, then her investment propensity would be 0.2 in that round.

In Figure 4, we report the average investment fraction across rounds by participants in each of these three treatment conditions. We are predominantly interested in the later rounds once investment behaviour has stabilised (see footnote 21 for a discussion of the reasons). The figure shows how over time investment behaviour between treatments diverges, converging towards the pattern of behaviour displayed in Figure 3.

---

<sup>21</sup>We focus on the later rounds in which: (i) the stakes for investing were higher, (ii) participants have familiarised themselves with the task, and (iii) accumulated a large database. The first ten rounds were intended to allow subjects to learn about how the game worked, and also to fill their databases with a sufficient number of observations to reduce the influence of inference from small samples. Figure 4 shows how the propensity to invest evolved over the twenty rounds in each treatment. Furthermore, the figure shows that the investment propensity appeared to have stabilised by the last five rounds.

Figure 4: Propensity to invest across rounds, by treatment

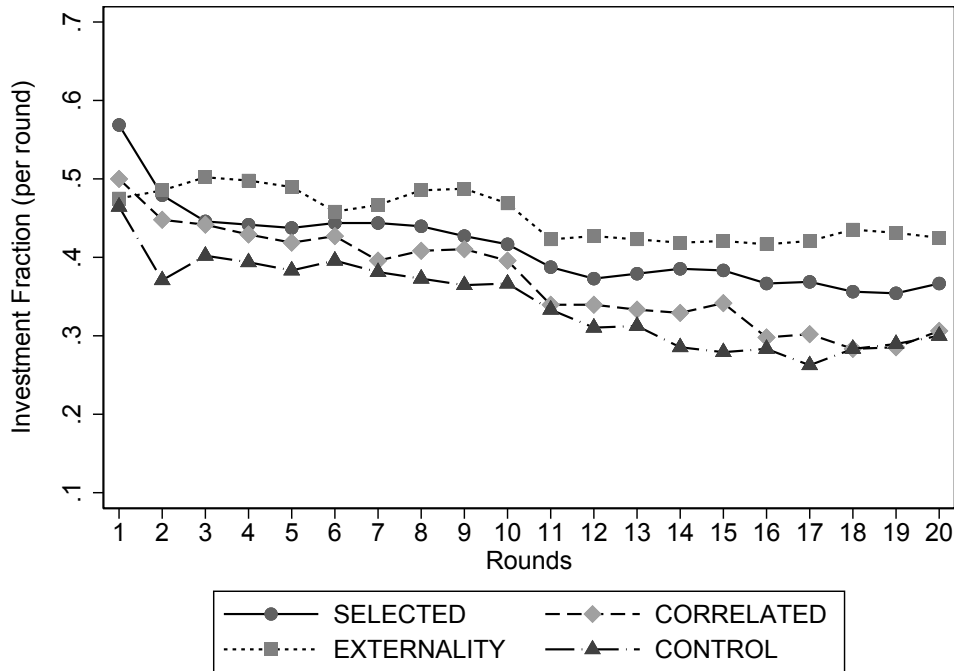


Table 2 tests whether there are statistically significant differences in the propensity to invest between treatments. Column (1a) reports treatments differences in the propensity to invest over all twenty rounds, while (1b) restricts attention to the last five rounds. As a robustness check, in column (2) we consider the restricted sample of individuals who followed a threshold strategy in each of the last five rounds.<sup>22</sup> Furthermore, in all regressions we adopt the conservative approach of including each individual as a single observation. The results in the three regressions are fairly consistent, with participants investing between 27% and 34% of the time in the control treatment, and an increase in investment by 7-8 pp in SELECTED. This provides our first result.

**Result 1.** *(Hypothesis 1) Individuals who are given access to a selected subset of past outcomes generated by others mistakenly increase their propensity to invest.*

<sup>22</sup>For example, an individual is removed from the restricted sample if she invested for an attribute value of 6, but not for an attribute value of 7, in one of the last five rounds.

Table 2: Propensity to invest by treatment

	Full Sample		Restricted Sample
	All (1a)	R16-20 (1b)	R16-20 (2)
SELECTED	0.07*** (0.03)	0.08*** (0.03)	0.07* (0.04)
EXTERNALITY	0.11*** (0.03)	0.14*** (0.03)	0.15*** (0.03)
CORRELATED	0.03 (0.03)	0.01 (0.03)	-0.03 (0.04)
Constant	0.34*** (0.02)	0.28*** (0.02)	0.27*** (0.02)
Observations	192	192	148
Adjusted $R^2$	0.070	0.113	0.178

Notes: (i) OLS regressions include one observation per individual, (ii) Dependent variable is individual average investment propensity, either over all rounds, or rounds 16-20. (iii) Standard errors in parentheses, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

### Do more informed individuals exert a negative externality on others?

We study this question by comparing SELECTED with EXTERNALITY. Recall that in EXTERNALITY, we exogenously vary the degree of selection by replacing one of the three human players in each group with an omniscient computer player, who only invests in successful projects. This is known to participants.

The information externality of introducing the computer players is illustrated by Figure 13 in the Appendices, which portrays the average empirical database observed in round 20. For every attribute value, there is an upward shift in the fraction of observed past projects that were successful when comparing the SELECTED and EXTERNALITY treatments.



Table 3: Propensity to invest in SELECTED and EXTERNALITY

	Full Sample		Restricted Sample
	All (1a)	R16-20 (1b)	R16-20 (2a)
EXTERNALITY	0.04 (0.03)	0.06** (0.03)	0.08** (0.03)
Constant	0.41*** (0.02)	0.36*** (0.02)	0.34*** (0.03)
Observations	96	96	73
Adjusted $R^2$	0.010	0.034	0.052

Notes: (i) OLS regressions include one observation per individual, (ii) Dependent variable is individual average investment propensity, either over all rounds, or rounds 16-20. (iii) Standard errors in parentheses, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 3 tests whether this information externality translates into a shift in investment behaviour. Indeed, the regression estimates suggest that the propensity to invest is 6 to 8 pp higher in the later rounds in EXTERNALITY in comparison to SELECTED.<sup>23</sup> This upward shift in investment is particularly noteworthy in light of the fact that participants are already overinvesting in the SELECTED treatment. This is reflected by the results in Table 2 which show that investment is 11 - 15pp higher in EXTERNALITY relative to CONTROL, representing a 30 - 50% increase.<sup>24</sup>

**Result 2.** (*Hypothesis 2*) *When the decision making of one individual is improved, this exerts a negative externality on those learning from her outcomes.*

### Is the influence of selection neglect ameliorated in situations where signals are correlated?

The discussion above has illustrated the harmful effects selection neglect can exert on decision making. These effects are driven by the failure to pay attention to the private information guiding

<sup>23</sup>We do not observe a significant difference in investment when we pool all twenty rounds. Examining Figure 4 shows that during the first ten periods, while participants face low incentives and are still learning and accumulating data, the gap between the two treatments is smaller than in the later periods when the stakes are higher.

<sup>24</sup> Furthermore, the introduction of computer players in EXTERNALITY may have increased the salience of the selection taking place, since the computers follow a clear, simple investment rule. An increase in the salience of selection could have lead to a reduction in investment. However, if the introduction of computer players did indeed increase the salience of the selection for some participants, the observed overall increase in investment in EXTERNALITY suggests that selection neglect was the dominant force.

a decision makers choices, implying a selected subset of outcomes is observed. However, it is important to notice that the influence of selection neglect varies in a systematic way according to the characteristics of the decision making context. In particular, in situations where different individuals focus on similar dimensions of the decision problem, the selection neglect should play a smaller role. Our CORRELATED treatment allows us to test whether increasing the correlation in the signals received by different individuals about a given project causes the harmful influence of selection neglect to dissipate.

Figures 3 and 4 suggest that by the later rounds, the propensity to invest in the CORRELATED treatment converges to the same level as the CONTROL treatment in agreement with the theoretical prediction. Table 2 supports this conclusion, showing that there are no significant differences in investment between the CORRELATED and CONTROL treatments. Furthermore, the point estimate of the difference between treatments is small, ranging from 3pp to -3pp across the three specifications.

**Result 3.** *(Hypothesis 3) An increase in the correlation in signals received by different individuals ameliorates the influence of selection neglect when learning from others.*

### **The manifestation of the treatment effect**

The evidence presented above focused on differences in the mean propensity to invest across all signals / attribute values that an individual observes before choosing to invest in a project. To complement this, one may look at the full investment strategies across treatments. For Selection Neglect subjects, investment threshold should be shifted downwards in the SELECTED and EXTERNALITY treatments — implying that the treatment effect should be concentrated at intermediate attribute values (e.g. an investor who would invest for all attribute values of 8 or higher in the CONTROL treatment, might, in the SELECTED treatment, instead invest for all attribute values 7 or higher, and in the EXTERNALITY treatment, invest for all attribute values 6 or higher).<sup>25</sup>

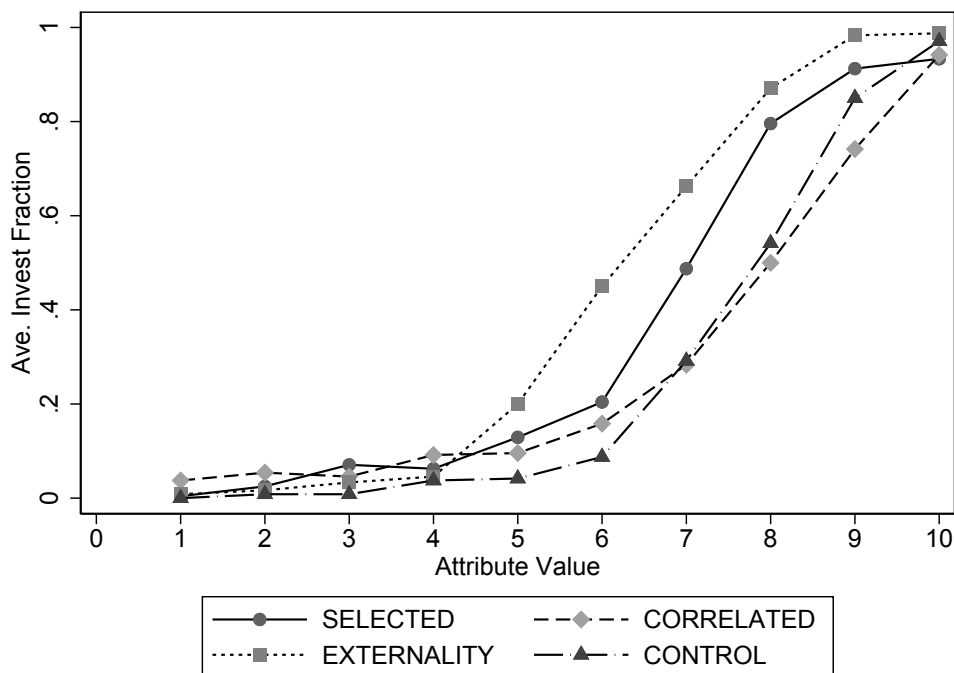
Figure 5 shows that our data is consistent with this, with the treatment differences arising at intermediate attribute values. The figure displays the average participant's propensity to invest, conditional on each attribute value in the last five rounds. Participants appear to have understood the task well as we observe very low investment rates for attributes 1 to 4. Similarly, we observe

---

<sup>25</sup>That is, in all treatments, low attribute values should be below an individual's threshold, and high attribute values be above their threshold. Therefore, individuals should never invest for these low attribute values, and always invest for high attribute values. A shift downwards in their threshold would increase the propensity to invest at intermediate attribute values.

very high investment rates for attributes 9 and 10. The differences in investment between treatment occurs predominantly between attributes 6 and 8.

Figure 5: Investment by attribute value, between treatments (R16-20)



## 5 Discussion

This paper studies how individuals learn from others' experience. In particular, we focus on how observing a partial database of outcomes can lead individuals to make mistakes in their inference and choices when displaying selection neglect. We consider four feedback structures, and show that behaviour across these structures reflects the pattern of behaviour expected of a selection neglect agent.

### 5.1 Heterogeneity and risk attitudes

Our data reveal that the chosen investment strategies are heterogeneous among subjects. It is tempting to attribute the heterogeneity to differences in risk attitudes.

Given that the observed propensity to invest in CONTROL and CORRELATED is *above* that implied by risk neutral agents in the Bayesian case, our investment data would suggest that subjects

exhibit some form of risk loving behavior. However, this is in contrast to the behaviour in the risk elicitation task that we observe which is in line with what is usually observed. If we take the risk elicitation data as a baseline, we would then have to conclude that the investment environment triggers some taste for investing, which is perhaps expected, since in an experiment about investments, subjects may want to invest.<sup>26</sup>

Regarding the observed heterogeneity, we are hence faced with a situation where subjects may vary in three fundamental dimensions, their risk attitude, their utility boost from investments as such and their degree of selection neglect. It is not possible to separately identify these three dimensions on the individual level. We do not view this as a deeper problem though, as the question about the precise joint distribution of these traits is orthogonal to the main questions we pose in this paper — whether there is selection neglect on average and whether its impact depends on both the sophistication of others and the correlation in the signal structure.

## 5.2 Inference from selected data

The existing literature studying belief formation from selected data is small, which is somewhat surprising given the proliferation of selected data in daily life. However, recent work has demonstrated how neglect of selected data can generate striking deviations from rational behaviour in contexts very different to those considered in the present paper. In an interesting experiment that builds on the theoretical work in [Esponda \(2008\)](#) and [Esponda and Pouzo \(2014\)](#), [Esponda and Vespa \(2018\)](#) consider a committee decision problem where the participant and two computer players decide jointly whether to invest in a risky project. The computers follow a noisy decision rule, governed by an unknown parameter. When the computer’s decisions are correlated with the true state of the world, the participant observes a selected sample of past projects. Their paper shares with the current paper its consideration of the neglect of selected data. However, it differs in terms of the strategic setting, presence of an experimentation motive (through learning from own experience), and requirement of hypothetical reasoning. In line with the results of this paper, [Esponda and Vespa \(2018\)](#) find evidence of neglect of endogenous selection.

In work complementary to the current paper, [Enke \(2017\)](#) studies whether individuals are able to take into account missing information. The paper presents a careful and insightful experimental

---

<sup>26</sup>Alternatively, the investment environment may instead trigger a shift in overconfidence (i.e. the investment environment generates an upward shift in beliefs). This would imply a higher propensity to invest in the investment environment. This perspective takes the view that there are different sources of overconfidence — (i) an upward shift in beliefs triggered by some environments, and (ii) an upward shift in beliefs generated by the selected data subjects observe in some contexts.

investigation into the existence of selection neglect in the context of individuals selecting themselves into news networks that reflect their own views (sometimes referred to as “echo chambers”). Enke (2017) shows that selection neglect could lead to the polarisation of beliefs and provides evidence on the cognitive mechanism underlying the bias. While exploring a similar theme, the focus on how selection neglect in “echo chambers” can lead to polarisation of views is different to the current paper’s focus on a micro-foundation of overconfidence. Interestingly, while Enke (2017) demonstrates that selection neglect can lead to polarisation of views, we demonstrate that in a large class of situations, the same bias can also result in all individuals holding beliefs that are biased in the same direction. This is indicative of the importance of the context in determining the predicted behaviour of selection neglect agents.

### 5.3 Overconfidence

Taken together, the nascent evidence discussed above suggests that selection neglect plays a significant role in influencing behaviour across a range of contexts. However, an important lesson from this literature is that, depending on the context, selection neglect can manifest itself as highly distinctive behaviours. In the contexts considered in the current paper, selection neglect manifests as behaviour akin to that generated by overoptimism or overconfidence provided that signals across different investors are not perfectly correlated. When they are, the apparent overconfidence disappears.

This distinction highlights that the overoptimism bias that we observe is neither the mere consequence of some general belief distortion nor simply due to the fact that the available information is biased towards more profitable projects as this would be true whether or not signals are correlated. Instead the overoptimism that we observe is affected by whether subjects look at the same aspect of a project or at different aspects when they make their judgments. Our study, hence, suggests that overoptimism should be expected to be more prevalent in complex investment problems than in simple ones. Similarly, on the population level, overoptimism may be more common in heterogeneous societies with competing education and value systems than in more uniform societies where people are trained in similar ways.

### 5.4 Survivorship bias and reference group neglect

Two concepts that are related to the ideas discussed in this paper are *survivorship bias* and *reference group neglect*. Survivorship bias typically refers to situations in which individuals draw inference from a distribution that is truncated to the left, normally because the observations in the left tail

have “failed” in some sense. For example, [Denrell \(2003\)](#) studies how false beliefs can result from an undersampling of failed ventures, and [Brown et al. \(1992\)](#) demonstrate that survivorship can induce an apparent persistence in returns in the performance of mutual funds. Since the concept of selection neglect that we consider shares some common features with survivorship bias, one might conjecture that individuals who are prone to selection neglect might also be prone to biased inference when observing the outcomes of a survival process. While this may be the case, there are several key differences between our concept of selection neglect and survivorship bias. Firstly, in contrast to a survival process, which may be viewed as an exogenous selection process, we consider an endogenous selection process, analysed using an equilibrium approach. This generates the externality of well informed agents that we test. Secondly, we study individuals who make the selection decision on the basis of private information, implying the selection is more subtle than in a survival process. Furthermore, this generates predictions regarding the variation in influence of selection neglect as a function of the information structure (e.g. the amelioration under highly correlated signals).

Reference group neglect is a concept discussed by [Camerer and Lovallo \(1999\)](#) in their seminal experimental study of whether excess entry and business failure can be explained by overconfidence. In particular, in their experiment excess entry is exacerbated when subjects self-select into participating in the experiment, presumably on the basis of believing that they are highly skilled in the relevant area. The authors attribute this exacerbation of the bias to the fact that subjects fail to adjust their entry decision to incorporate the self-selection of their competitors. Two key differences between reference group neglect and the selection neglect considered in this paper are: (i) that [Camerer and Lovallo \(1999\)](#) consider a process where individuals, rather than projects are selected, which drives the pattern of predicted behaviour across different information structures in our paper;<sup>27</sup> (ii) the current paper focuses on situations where individuals learn from the outcomes of others, which is different in spirit to the neglect of self-selection considered by [Camerer and Lovallo \(1999\)](#).

## 6 Conclusion

The information revolution of recent decades has led to an explosion of easily available, large datasets. In particular, we are increasingly able to observe both objective and subjective outcomes

---

<sup>27</sup>Interestingly, when one considers the neglect of self-selection of people, there are also situations in which the predictions of this neglect can generate pessimism rather than optimism (e.g. when learning from the outcomes associated with a harmful decision, such as smoking or drug usage).

(e.g. online reviews<sup>28</sup>) of other people’s choices relating to a wide range of life decisions. This information can help us to make informed decisions, by learning from the mistakes and successes of others. However, it comes with the challenge of drawing accurate inference from the data we observe. In particular, in many cases, the outcome of interest is only observed for a specific subset of individuals. It is therefore imperative that we understand how individuals learn from this type of selected data, and are aware of the types of mistakes it can generate.

This paper presents the results from an experimental investment game that captures the spirit and core mechanisms discussed in the theoretical exploration of selection neglect in [Jehiel \(2018\)](#). We test and find strong support for the validity of the assumption that people display a neglect of the way in which data observed in their environment is selected. Additionally, we run treatments to illustrate the influence of this bias on behaviour under different benchmark information structures. We show that specific characteristics of the decision making context can be used to systematically predict whether the influence of selection bias is exacerbated or ameliorated. Taken together, our results across information structures are fully consistent with the theoretical predictions for the behaviour one would expect when subjects neglect the influence of selection.

Our results have important implications because the core elements of our decision problem relate to many of the biggest decisions we make in our lives (e.g. about one’s career, whether and where to buy a house, where to attend college, whether to switch jobs, whether to get married, and whether to have children). The results from this paper, and the related literature, point towards a need for ensuring that decision makers are made aware of the way in which the data they observe has been selected, and the detrimental influence this can have on their decision making if not accounted for.

It seems that while we are all “everyday econometricians” in our daily lives, like econometricians of the past, from time to time we fall into the trap of neglecting the influence of selection on the data we observe; unfortunately this leads to biased inference and decision making.

---

<sup>28</sup>Note, online reviews are only observed for the subset of individuals who chose (i) to buy the product, or participate in the activity, and then (ii) to write a review.

## References

- Bénabou, R. and J. Tirole (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives* 30(3), 141–64.
- Brown, S. J., W. Goetzmann, R. G. Ibbotson, and S. A. Ross (1992). Survivorship bias in performance studies. *The Review of Financial Studies* 5(4), 553–580.
- Camerer, C. and D. Lovo (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review* 89(1), 306–318.
- Chen, D. L., M. Schonger, and C. Wickens (2016). otree: An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance* 9, 88–97.
- Compte, O. and A. Postlewaite (2004). Confidence-enhanced performance. *American Economic Review* 94(5), 1536–1557.
- Cooper, A. C., C. Y. Woo, and W. C. Dunkelberg (1988). Entrepreneurs' perceived chances for success. *Journal of Business Venturing* 3(2), 97–108.
- Dawson, C., D. de Meza, A. Henley, and G. R. Arabsheibani (2014). Entrepreneurship: cause and consequence of financial optimism. *Journal of Economics & Management Strategy* 23(4), 717–742.
- De Bondt, W. F. and R. H. Thaler (1995). Financial decision-making in markets and firms: A behavioral perspective. *Handbooks in operations research and management science* 9, 385–410.
- Denrell, J. (2003). Vicarious learning, undersampling of failure, and the myths of management. *Organization Science* 14(3), 227–243.
- Eil, D. and J. Rao (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics* 3(2), 114–138.
- Enke, B. (2017). What you see is all there is. *Working Paper*.
- Enke, B. and F. Zimmermann (2017). Correlation neglect in belief formation. *Working Paper*.
- Esponda, I. (2008). Behavioral equilibrium in economies with adverse selection. *American Economic Review* 98(4), 1269–1291.



- Esponda, I. and D. Pouzo (2014). Conditional retrospective voting in large elections. *Working Paper*.
- Esponda, I. and E. Vespa (2018). Endogenous sample selection: A laboratory study. *Quantitative Economics* 9(1), 183–216.
- Eyster, E. and M. Rabin (2005). Cursed equilibrium. *Econometrica* 73(5), 1623–1672.
- Graeber, T. (2018). Inattentive inference. *Working Paper*.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association* 1(1), 114–125.
- Heckman, J. (1990). Varieties of selection bias. *American Economic Review* 80(2), 313–318.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*.
- Jehiel, P. (2005). Analogy-based expectation equilibrium. *Journal of Economic Theory* 123(2), 81–104.
- Jehiel, P. (2018). Investment strategy and selection bias: An equilibrium perspective on overoptimism. *American Economic Review*. Forthcoming publication.
- Jehiel, P. and F. Koessler (2008). Revisiting games of incomplete information with analogy-based expectations. *Games and Economic Behavior* 62(2), 533–557.
- Jin, G. Z., M. Luca, and D. Martin (2018). Is no news (perceived as) bad news? an experimental investigation of information disclosure. *Working Paper*.
- Koehler, J. J. and M. Mercer (2009). Selection neglect in mutual fund advertisements. *Management Science* 55(7), 1107–1121.
- Koellinger, P., M. Minniti, and C. Schade (2007). ‘i think i can, i think i can’: Overconfidence and entrepreneurial behavior. *Journal of Economic Psychology*.
- Logg, J., U. Haran, and D. A. Moore (2018). Is overconfidence a motivated bias? experimental evidence. *Working Paper*.
- Malmendier, U. and G. Tate (2005). Does overconfidence affect corporate investment? CEO overconfidence measures revisited. *European Financial Management* 11(5), 649–659.

- Malmendier, U. and T. Taylor (2015). On the verges of overconfidence. *Journal of Economic Perspectives* 29(4), 3–8.
- Möbius, M., M. Niederle, P. Niehaus, and T. Rosenblat (2014). Managing self-confidence. *Working Paper*.
- Moore, D. A. and P. J. Healy (2008). The trouble with overconfidence. *Psychological review* 115(2), 502.
- Puri, M. and D. T. Robinson (2007). Optimism and economic choice. *Journal of Financial Economics* 86(1), 71–99.
- Rabin, M. (2002). Inference by believers in the law of small numbers. *Quarterly Journal of Economics* 117(3), 775–816.
- Schwardmann, P. and J. J. Van der Weele (2016). Deception and self-deception. *Working Paper*.
- Selten, R. (1967). Die strategiemethode zur erforschung des eingeschränkt rationalen verhaltens im rahmen eines oligopolexperiments. In H. Sauer mann (Ed.), *Beiträge zur experimentellen Wirtschaftsforschung*, pp. 136–168. Tübingen: Mohr.
- Spiegler, R. (2011). *Bounded rationality and industrial organization*. Oxford University Press.
- Tversky, A. and D. Kahneman (1971). Belief in the law of small numbers. *Psychological Bulletin*.
- Von Hippel, W. and R. Trivers (2011). Reflections on self-deception. *Behavioral and Brain Sciences* 34(1), 41–56.
- Zimmermann, F. (2018). The dynamics of motivated beliefs. *Working Paper*.

# APPENDICES

## Appendix A: Theory — some further details

### A.1 Calculating the equilibrium in scenario 1

As noted above, without loss of generality, due to the symmetry of the game we can consider the perspective of a type A agent. Making use of the two observations that: (i)  $P(\bar{x}|s_A; s_B \geq s^{SN1}) = \sum_{s_B} P(\bar{x}|s_A; s_B)$ , and (ii)  $P(\bar{x}|s_A; s_B) = \frac{s_A + s_B - 11}{10}$  if  $s_A + s_B \geq 12$  and 0 otherwise, equation 2 can be expanded using the following analytical expression:

$$\begin{aligned} \widehat{P}(\bar{x}|s_A; q_{inv}^{SN1}) &= P(\bar{x}|s_A; s_B \geq s^{SN1}) \\ &= \frac{(\widetilde{s^{SN1}} - 11) \cdot (2s_A + \widetilde{s^{SN1}} - 12)}{20 \cdot (s^{SN1} - 11)} \end{aligned} \quad (3)$$

where

$$\widetilde{s^{SN1}} := \begin{cases} s^{SN1} & \text{if } s_A + s^{SN1} \geq 12 \\ 12 - s_A & \text{if } s_A + s^{SN1} < 12 \end{cases}$$

This allows us to construct a table consisting of the expected fraction of successful projects in a type A agent's feedback, conditional on signal  $s_A$  and the threshold strategy being used by the other agents,  $s^{SN1}$  (see Table 4 below). Therefore, each column of the table reports the success fractions that a type A agent would observe in her feedback if all other players were following a specific strategy  $s^{SN1}$ , as the number of projects in her feedback gets large.

To illustrate this, let us consider how the feedback of a type A agent changes, depending on the strategy followed by the agents generating her feedback. The right-most column shows the success fractions observed if the feedback is generated by players who only invest after receiving a signal of 10. For a type A agent with this type of feedback, on average 50% of the projects that she observes with  $s_A = 6$  are successful projects. In contrast, the left-most column shows the success fractions observed if the feedback is generated by players who invest for a signal of 1 or higher (i.e. they always invest). For a type A agent with this type of feedback, on average 15% of the projects that she observes with  $s_A = 6$  are successful projects. The success fractions in this left-most column coincide with the Bayesian success probabilities (since there is no selection in

the feedback).

Table 4: Expected success fractions for observed projects

	$s^{SN1} =$									
	1	2	3	4	5	6	7	8	9	10
$s_A =$										
1	0	0	0	0	0	0	0	0	0	0
2	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.03	0.05	0.1
3	0.03	0.03	0.04	0.04	0.05	0.06	0.08	0.1	0.15	0.2
4	0.06	0.07	0.08	0.09	0.1	0.12	0.15	0.2	0.25	0.3
5	0.1	0.11	0.13	0.14	0.17	0.2	0.25	0.3	0.35	0.4
6	0.15	0.17	0.19	0.21	0.25	<b>0.3</b>	0.35	0.4	0.45	0.5
7	0.21	0.23	0.26	0.3	0.35	0.4	0.45	0.5	0.55	0.6
8	0.28	0.31	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7
9	<b>0.36</b>	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8
10	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9

Bayesian agents therefore use the probabilities in the left-most column and invest for all signals  $s_A$  such that  $P(\bar{x}|s_A) > \frac{c}{\bar{x}} \approx 0.294$ . Clearly, this inequality is only satisfied for  $s_A = 9$  and  $s_A = 10$ , implying that the risk neutral Bayesian agent will follow the threshold strategy  $s^{Bayes} = 9$ .

To find the symmetric equilibrium investment strategy for coarse investors, recall from the main text that the symmetric pure strategy equilibrium is given by  $s^{SN1}$  should be such that  $\hat{P}(\bar{x}|s^{SN1}; q_{inv}^{SN1}) > \frac{c}{\bar{x}}$  and  $\hat{P}(\bar{x}|s^{SN1} - 1; q_{inv}^{SN1}) < \frac{c}{\bar{x}}$  if  $s^{SN1} > 1$ . Since one can read the beliefs of a coarse investor,  $\hat{P}(\bar{x}|s_A; q_{inv}^{SN1}) = P(\bar{x}|s_A; s_B \geq s^{SN1})$ , directly off Table 4, one can see that the symmetric equilibrium in pure strategies is given by  $s^{SN1} = 6$ . The logic is as follows: given that  $\frac{c}{\bar{x}} = \frac{1}{3.4} \approx 0.294$ , moving down the  $s^{SN1} = 6$  column in Table 4 shows that the Type A individual who holds these beliefs will invest for  $s_A = 6$ , but not for  $s_A = 5$ .

The information contained in Table 4 can be depicted graphically as in Figure 6. This visualization of the data provides a second illustration of the equilibrium threshold strategies followed by the Bayesian and coarse investor. In Figure 6, the Bayesian investor's beliefs are represented by the lowest (blue) curve.<sup>29</sup> The ratio  $\frac{c}{\bar{x}} \approx 0.294$  is drawn as a horizontal (light blue) line. The Bayesian agent invests for signals which yield a belief about the probability of success that exceeds the ratio  $\frac{c}{\bar{x}}$  (i.e. when the dark blue curve is above the light blue horizontal line). Clearly this is the case for

<sup>29</sup>This lowest (blue) curve depicts the information contained in the left-most column of Table 4.

signals  $s_A = 9$  and  $s_A = 10$ .

For the coarse investor, each of the curves in Figure 6 depict her beliefs as a function of the threshold strategy being followed by the individuals generating her feedback. For example, the top-most curve reflects the coarse investors beliefs when all other agents are following a threshold strategy, where they only invest after a signal of 10, but not for lower signals. Looking at the  $s^{SN1} = 6$  curve shows that the individual will invest for  $s_A = 6$ , but not for  $s_A = 5$  (since the first belief is above the horizontal line, while the second is below).

Figure 6: Selection Neglect Agent's Perceived Success Probabilities  
by Threshold Strategy of Others

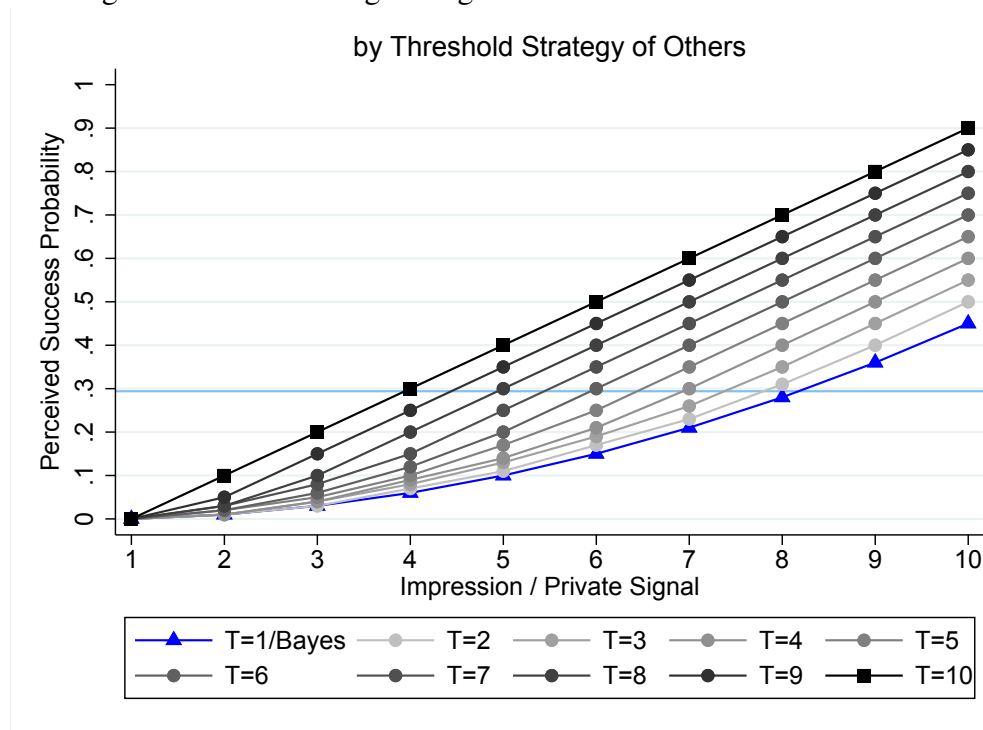


Figure 6 also provides an illustration of how individuals can exert an externality on others through the feedback generated. As the threshold strategy followed by other agents moves upwards, the coarse agent's beliefs become increasingly distorted. This leads to poorer decision making by the coarse agent. For example, if other agents are following the rational threshold strategy,  $s^{Bayes} = 9$ , the coarse agent will hold highly distorted beliefs (i.e., refer to the distance between the T=1/ Bayes curve and the T=9 curve).

## A.2 Calculating the equilibrium in scenario 2

As discussed in the main text above, in scenario 2, for all  $s_A \geq s^{SN2}$ ,  $\widehat{P}(\bar{x}|s_A; q_{inv}^{SN2}) = P(\bar{x}|s_A)$ . The feedback received when there are perfectly correlated signals symmetric play is shown in Table 5. However, allowing for some trembling would imply that every column of the table is equivalent to the left-most column which contains the Bayesian success probabilities, conditional on  $s_A$ . Therefore, in this scenario risk neutral coarse investor chooses  $s^{SN2} = s^{Bayes}$ .

Table 5: Expected success fractions for observed projects

	$s^{SN2} =$									
	1	2	3	4	5	6	7	8	9	10
$s_A =$										
1	0									
2	0.01	0.01								
3	0.03	0.03	0.03							
4	0.06	0.06	0.06	0.06						
5	0.1	0.1	0.1	0.1	0.1					
6	0.15	0.15	0.15	0.15	0.15	0.15				
7	0.21	0.21	0.21	0.21	0.21	0.21	0.21			
8	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28		
9	<b>0.36</b>	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	<b>0.36</b>
10	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45

### A.3 Calculating the equilibrium in scenario 3

Recognising that the omniscient players invest iff a project is successful gives  $\mathbb{E}_{s_B, s_C}[q^R(s_A, s_B, s_C)] = P(\bar{x}|s_A)$ , which implies that:

$$\begin{aligned}
 \widehat{P}(\bar{x}|s_A; q_{inv}^{SN3}) &= \frac{\sum_{s_B} [(1 - \lambda) \cdot P(\bar{x}|s_A; s_B) \cdot 1_{s_B \geq s^{SN3}} + \lambda \cdot P(\bar{x}|s_A)]}{\sum_{s_B} [(1 - \lambda) \cdot 1_{s_B \geq s^{SN3}} + \lambda \cdot P(\bar{x}|s_A)]} \\
 &= \frac{(1 - \lambda) \cdot \widehat{P}(\bar{x}|s_A; q_{inv}^{SN1} = q_{inv}^{SN3}) \cdot (11 - s^{SN3}) + \lambda \cdot P(\bar{x}|s_A) \cdot 10}{(1 - \lambda) \cdot (11 - s^{SN3}) + \lambda \cdot P(\bar{x}|s_A) \cdot 10} \\
 &= \frac{(1 - \lambda)[(11 - \widetilde{s^{SN3}}) \cdot (s_A - 11) + \frac{1}{2}(10 \cdot 11 - \widetilde{s^{SN3}}(s^{SN3} - 1))] + \lambda \cdot \frac{s_A^2 - s_A}{2}}{(1 - \lambda) \cdot 10 \cdot (11 - s^{SN3}) + \lambda \cdot \frac{s_A^2 - s_A}{2}}
 \end{aligned} \tag{4}$$

where

$$\widetilde{s^{SN3}} := \begin{cases} s^{SN3} & \text{if } s_A + s^{SN3} \geq 12 \\ 12 - s_A & \text{if } s_A + s^{SN3} < 12 \end{cases}$$

Using the parameterisation relevant for our experimental design (i.e.  $\lambda = \frac{1}{2}$ ) yields the following table of beliefs of a type A coarse investor as a function of the threshold followed by other non-omniscient players,  $s^{SN3}$ , and her own signal,  $s_A$ .

Table 6: Expected success fractions for observed projects

		$s^{SN3} =$									
		1	2	3	4	5	6	7	8	9	10
$s_A =$											
1		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2		0.02	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.10	0.18
3		0.06	0.06	0.07	0.08	0.10	0.11	0.14	0.18	0.26	0.38
4		0.11	0.13	0.14	0.16	0.18	0.21	0.26	0.33	0.42	0.56
5		0.18	0.20	0.22	0.25	<b>0.29</b>	<b>0.33</b>	0.40	0.48	0.57	0.70
6		0.26	0.29	0.32	0.35	0.40	0.46	0.53	0.60	0.69	0.80
7		0.35	0.38	0.42	0.46	0.52	0.58	0.64	0.71	0.78	0.87
8		0.44	0.47	0.52	0.57	0.63	0.68	0.74	0.79	0.85	0.92
9		0.53	0.57	0.62	0.67	0.72	0.77	0.82	0.86	0.91	0.96
10		0.62	0.67	0.71	0.76	0.80	0.84	0.88	0.92	0.95	0.98

For the Bayesian agent, the information in Table 6 is irrelevant. Instead she relies on her knowledge of the DGP and, as above, uses the probabilities  $P(\bar{x}|s_A)$  to guide her decision making (the left-most column of Table 4 contains these probabilities). She therefore again follows the threshold strategy  $s^{Bayes} = 9$ .

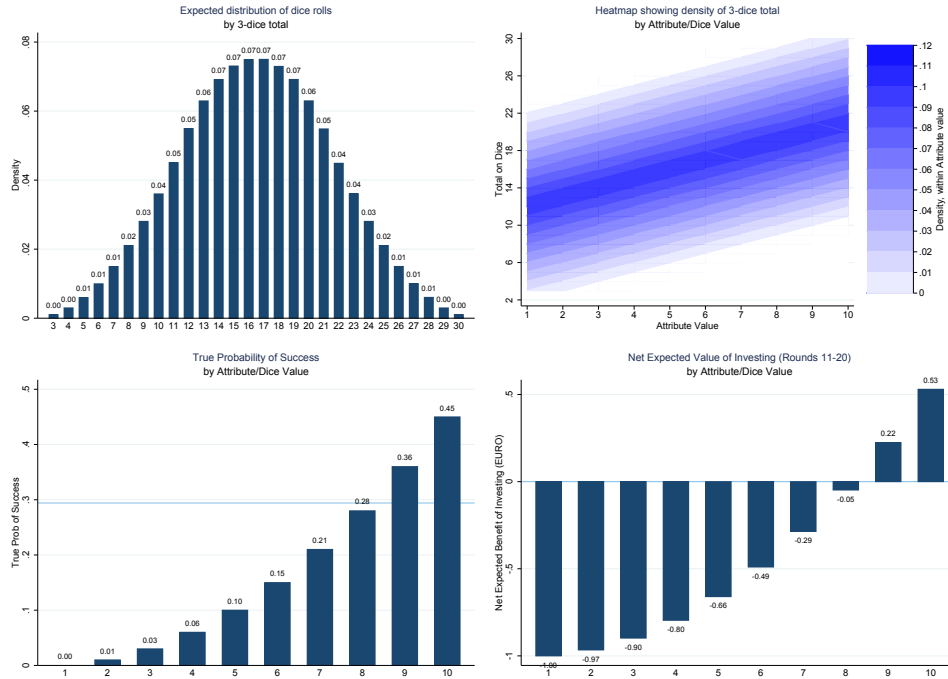
There is no pure strategy symmetric equilibrium for risk neutral coarse investors. Rather, there is exists a symmetric equilibrium in mixed strategies in which players mix between  $s^{SN3} = 5$  and  $s^{SN3} = 6$ , such that all players are indifferent between these two strategies.<sup>30</sup> This condition is satisfied when players choose  $s^{SN3} = 5$  with probability  $\mu = 0.8$ .

<sup>30</sup>This is the case when agents are indifferent between investing and not investing after the signal  $s_A = 5$  (i.e. when the perceived probability of success is 0.294118). The equilibrium mixture probability is not obtained by simply weighting the two relevant probabilities reported in Table 6, but rather by calculating the equilibrium, allowing for mixing.



## Appendix B: The data generating process (DGP)

Figure 7: Overview of the DGP and Decision Problem



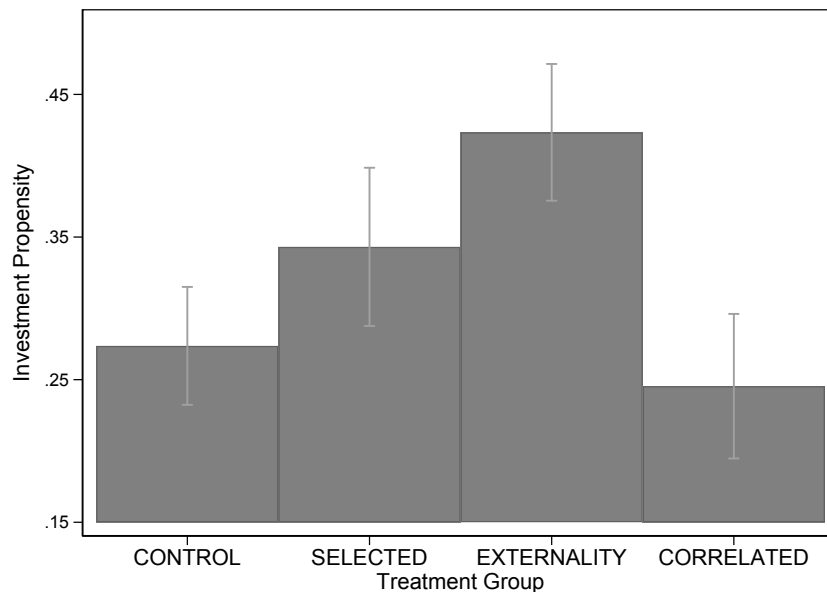
The data generating process is summarised in Figure 7.<sup>31</sup> The top left panel shows the unconditional distribution over the total when rolling three fair ten-sided dice. The top-right panel uses a heatmap to show the conditional distributions of the 3-dice totals, conditional on observing one of the dice values. Since the investor does not really care about the the full distribution of 3-dice totals, but rather cares about how this distribution maps onto the binary success / failure variable, the bottom-left panel displays the probability of success after observing each dice roll from 1 to 10. Given the parameters that we chose in our experiment in the last 10 rounds (i.e. the cost of investing, €1, and value of a successful investment, €3.40), the bottom-right panel then translates this into the net expected value of investing after observing each dice value from 1 to 10. This last panel illustrates that for the rational, risk neutral investor, it is only attractive to invest after observing a value of 9 or 10, although a value of 8 is marginal. A risk averse investor should be even more cautious about investing.

<sup>31</sup>The figures in the top two panels are generated by simulating 10 million projects and are therefore (fairly precise) approximations of the true distribution. The bottom two panels reflect the precise distribution.

## Appendix C: Robustness check

As a robustness check to our main results<sup>32</sup>, we also replicate our analysis for the subset of individuals who are “well behaved” in the sense that they follow a threshold strategy in each of the last five rounds. Restricting attention to this subsample rules out individuals who deviated from following a threshold strategy at least once (e.g. within a given round, invested for an attribute value of  $s$ , but did not invest for  $s'$  where  $s < s'$ ). This restriction is quite conservative as applies not only to individuals who, (i) made a mistake, or (ii) lacked understanding, but also to those who (iii) understood fully, but simply believed that a lower attribute had a higher probability of leading to a successful project in at least one instance. It is therefore reassuring that the vast majority of subjects (i.e. 77%) satisfy this restriction. We refer to this subsample as the “Restricted Sample”, both in the regression tables in the main text, and in the figures below. Restricting attention to this subsample does not change any of the results presented above. An additional benefit of focusing on “well behaved” subjects, who follow a threshold strategy, is that we can also present some results relating to the threshold strategy chosen. However, since an individual’s threshold strategy is mechanically related to her investment propensity, these results do not yield novel insights into behaviour.

Figure 8: Investment fraction in last five rounds, by treatment (Restricted Sample)



<sup>32</sup>The main results don't exclude any participants.

Figure 9: Propensity to invest across rounds, by treatment (Restricted Sample)

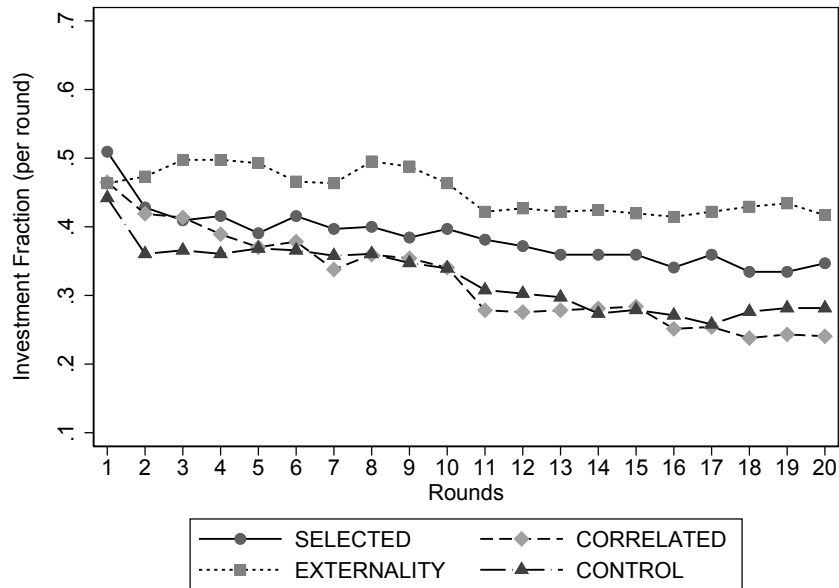


Figure 8 and 9 reproduce figures 3 and 4 respectively, for the restricted sample. The observed pattern of behaviour is very similar in the restricted sample to that observed in the full sample. This is supported by the regression results reported in tables 2 and 3, showing similar results for the restricted and full sample. Together, these results suggest that our treatment effects are not driven by individuals who failed to understand the game, or chose their strategy randomly.

Figure 10 shows the average threshold strategy followed across the twenty rounds for the restricted sample. We only present this figure for the restricted sample, as one needs to make further assumptions to assign a threshold to individuals who don't follow a proper threshold strategy. However, in our experimental design, the threshold strategy is the mechanical inverse of the investment propensity for individuals who follow a threshold strategy. Figure 10 uses the average of all individuals in the restricted sample for the last five rounds, but in the first fifteen rounds, additionally, for each round, excludes individuals who didn't follow a threshold strategy in that particular round. This explains why the figure 10 is not a perfect reflected image of figure 9 for the first fifteen rounds.

Figure 10: Average threshold strategy across rounds, by treatment (Restricted Sample)

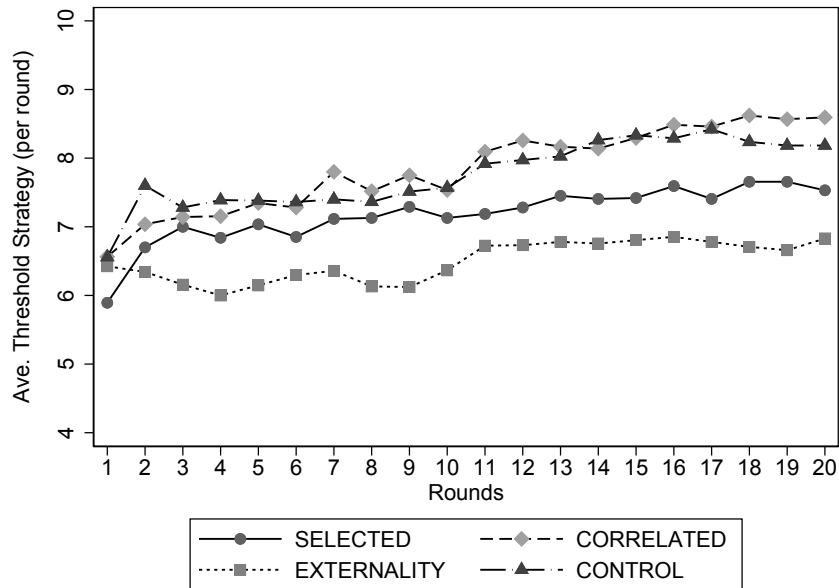


Figure 11 shows the average propensity to invest for each attribute value for the restricted sample. It is very similar to figure 5, with the exception that individuals in the restricted sample seem even better at not investing for attribute values 1 to 4 (as one might expect). Figure 12 provides some evidence on heterogeneity in the threshold strategy that participants followed in each of the treatments, but reporting the distribution of thresholds observed in round 20. In the figure, subjects are grouped according to the attribute value,  $s^*$ , such that they invested for all  $s \geq s^*$  in round 20. Therefore, participants with a threshold of  $s^* = 11$  did not invest for any attribute value. The figure shows that the treatment shifted the threshold distribution to the left in the SELECTED and EXTERNALITY treatments, relative to the CONTROL and CORRELATED treatments.

Figure 11: Investment by attribute value, between treatments (R16-20, restricted sample)

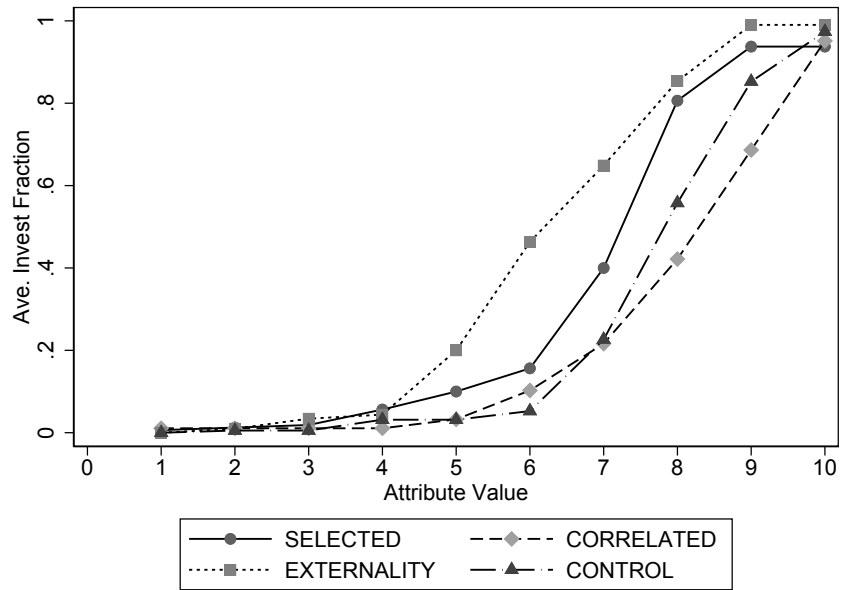
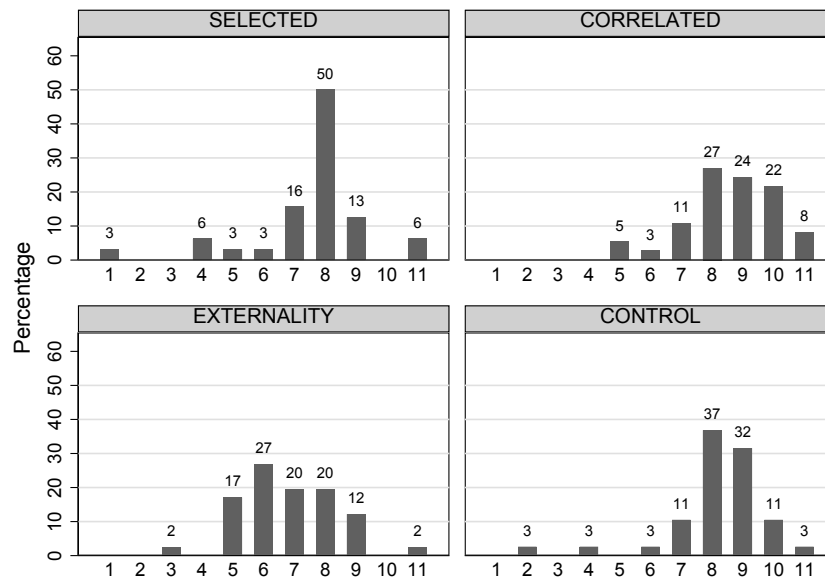
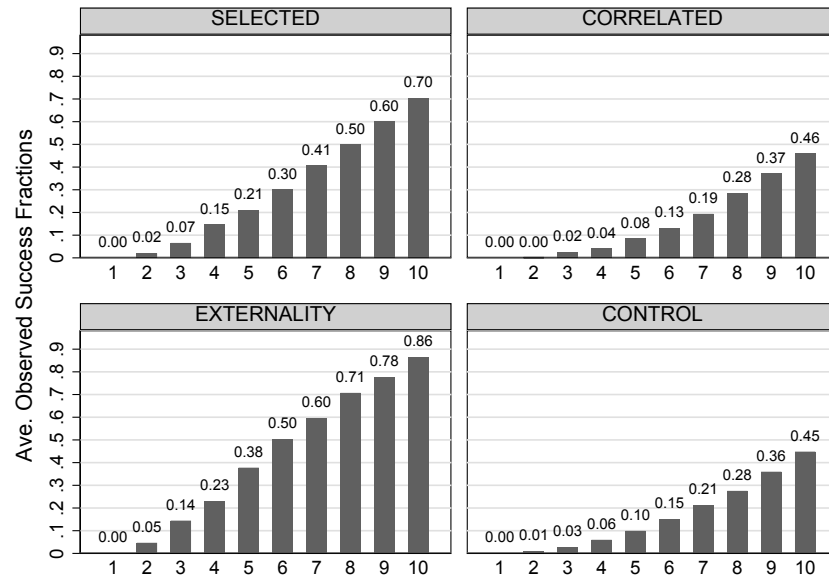


Figure 12: Distribution of threshold strategies in round 20 (restricted sample)



## Appendix D: Additional Figures and Tables

Figure 13: Average database observed by participants in round 20



## Appendix E: Instructions for the SELECTED treatment

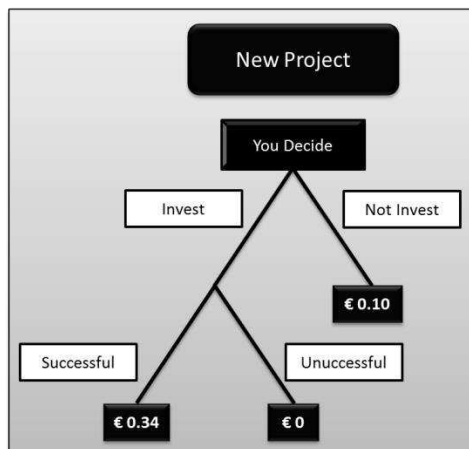
In Game I, you are going to have the opportunity to make a series of investment decisions. For each decision, you will be presented with a “project”. Every project will turn out to either be a *successful project* or an *unsuccessful project*. Each project will have three “attributes”. These attributes will be related to whether the project will be successful or not. You will be able to observe one of the three attributes for every project you face. (This will always be the same attribute).

In order to help you learn about whether a project you face will be a *successful project* or an *unsuccessful project*, you will be placed in a group with two other participants and you will be able to learn from the success or failure of the past investments of your group members. In particular, you will be able to observe whether projects similar to the one that you are currently considering were successful or not. Projects are “similar” when the attribute you observe is exactly the same between the projects. (We will provide you with more details on the exact information you will receive below).

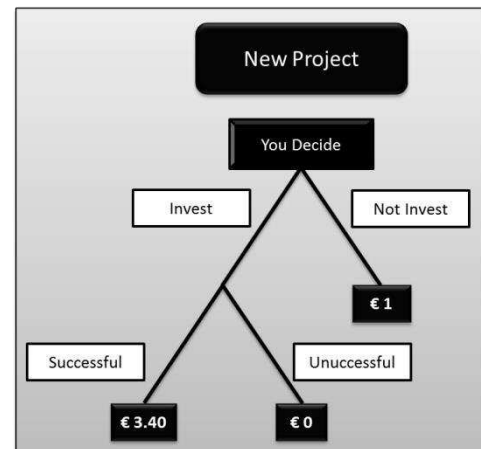
There will be two phases in the Investment Game. In Phase 1 (Low Stakes), each time you face a new project, you will be given an amount of money, €0.10. This is exactly the amount that it costs to make the investment. You can then choose whether to INVEST this €0.10; or NOT INVEST and keep the €0.10 to be paid at the end of the experiment. If you INVEST and the project is *successful*, you will be paid a high prize, which has value €0.34. If you INVEST and the project is *unsuccessful*, you will receive nothing and you will lose the €0.10 that you invested.

Phase 2 (High Stakes) is exactly the same, except all the amounts are multiplied by ten. This explanation is summarized in the following diagram.

**Phase 1: Low Stakes**



**Phase 2: High Stakes**



In **Phase 1 (Low Stakes)**, you will face low cost, low prize investment opportunities. In this phase, the cost of investment will be **€0.10** and the prize when the investment is successful will be **€0.34**. This phase will allow you to learn how the investment game works, and also to learn about which projects are likely to be *successful projects* and which projects are likely to be *unsuccessful projects*.

In **Phase 2 (High Stakes)**, the game is exactly the same. In particular, the chances that a project will be *successful* or *unsuccessful* are exactly the same as in Phase 1. The only difference is that the cost of investment will now be **€1.00** and the prize when the investment is successful will be **€3.40**.

#### Phase 1 (Low Stakes) investment costs and returns

Decision	Successful Project	Unsuccessful Project
Invest	Receive €0.34	Receive €0
Do Not Invest	Receive €0.10	

#### Phase 2 (High Stakes) investment costs and returns

Decision	Successful Project	Unsuccessful Project
Invest	Receive €3.40	Receive €0
Do Not Invest	Receive €1.00	

### *Learning about the project:*

In order to help you learn about whether a project will be *successful* or *unsuccessful*, you will be able to observe whether similar projects that were invested in by your group members in the past were successful. The text below provides you with details explaining: (i) what is meant by “similar” projects; and (ii) which past projects you will observe.

#### **Who is in your group?**

As mentioned above, you will be randomly assigned into a group with two other participants in this experiment (three in total, including yourself – called Group Members **A**, **B** and **C**). You will stay in this group for both Phase 1 and Phase 2 of the experiment.

#### **Which past projects will you observe?**

In order to learn about which projects are likely to be successful, you will be able to observe the success of projects that were invested in by the other members of your group. You will also observe one of the attributes of these past projects.

#### **What are project attributes?**

Every project has three attributes, called **a**, **b** and **c**. For every project, each of these three attributes takes an integer<sup>1</sup> value between 1 and 10. The attribute values are determined by the computer rolling three ten-sided dice – i.e. Attribute **a** is equal to the number shown on the purple dice, called Dice **a**; Attribute **b** is equal to the number shown on the red dice, called Dice **b**; and Attribute **c** is equal to the number shown on the green dice, Dice **c**. All three dice are fair dice (i.e. each dice has an equal chance of showing every number between 1 and 10).

<sup>1</sup> An integer is a whole number, so each attribute takes one of the following ten values: 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.



### ***Who observes which attributes?***

Each of the three group members (including you) will observe only one of the three attributes – and each group member will observe a different attribute from the other two.

More specifically, Group Member **A** will always observe Attribute **a**, Group Member **B** will always observe Attribute **b**; and Group Member **C** will always observe Attribute **c**.

### ***When is a project successful?***

The success of a project is determined by adding up the three attribute values (i.e. adding up the numbers on the three dice). If this total is equal to 22 or more (i.e. 22 to 30) then the project is *successful*; if the total is equal to 21 or less (i.e. 3 to 21) then the project is *unsuccessful*.

### ***Example: What information does Group Member C observe?***

#### *Information about current project that Group Member C is considering*

If, for example, you are Group Member **C**, then when you are considering a new project, you will always be able to observe Attribute **c** for this new project before you decide whether to invest in it.

#### *Information about past projects that Group Members A and B invested in*

As Group Member **C**, you would also have access to information about the Attribute **c** value of all projects that other members of your group invested in in the past. You will also be told whether these past projects were successful or not. In other words, you will know what proportion of the projects that other members of your group invested in with a particular Attribute **c** value were successful.

#### *An example*

Let's consider a concrete example to clarify this. Consider the following hypothetical project:

<b>Attribute a</b>	<b>Attribute b</b>	<b>Attribute c</b>	<b>Successful / Unsuccessful</b>
10	5	7	<b>Successful</b>

The information above is a complete summary of the project - it contains all three attribute values and the information about its success.

Recall that the three attribute values are determined by throwing three fair ten-sided dice (Dice **a**, Dice **b**, and Dice **c**). Furthermore, notice that the project is “successful” because adding up the three attribute values =  $10 + 5 + 7 = 22$ , which is between 22 and 30. Of course, as explained above, a participant in the experiment cannot observe such a complete description of projects.

Rather, assume that Group Member **C** is the one that has the opportunity to invest in this project. When she is deciding whether or not to invest, she will only observe the value of Attribute **c**:

Attribute a	Attribute b	Attribute c	Successful / Unsuccessful
		7	

In addition, Group Member **C** will also observe whether other past projects that (i) had a value of 7 for Attribute **c** and (ii) that either **A** or **B** invested in in the past; were successful or not. (You will see more detailed information about this below.)

*Now, what information will be revealed after **C** makes her decision?*

If she **decides not to invest**, then nobody, neither **A**, **B** nor **C** will receive any further information about this project.

If she **decides to invest**, then **after** the investment has been made, Group Member **A** will observe the following data about the project:

Attribute a	Attribute b	Attribute c	Successful / Unsuccessful
10			Successful

while Group Member **B** will observe:

Attribute a	Attribute b	Attribute c	Successful / Unsuccessful
	5		Successful

Notice that, while Group Members **A** and **B** observe whether the project that Group Member **C** invested in is successful or not, **C** herself will only find out at the end of the experiment. Group Members will never receive immediate feedback on the success of their own projects.

### **Summary**

The following summarises some of the key information from above:

- The game is completely symmetric for the three players (i.e. there is no difference between being assigned to be Group Member **A**, **B** or **C** at the start of the game. The only difference between players is the information they receive about the other two group members' past investments).
- Each of the three participants in your group will observe the value of one of the three attributes for: (i) all the projects they face, as well as (ii) all projects that other members of their group have previously invested in. Each will observe a different attribute.
- You will only observe the outcome of projects that your group members invested in.
- You won't observe the outcome of the projects that you invest in until the end of the experiment.
- The relationship between an attribute value and the chances of the corresponding project's success will stay the same throughout the experiment.
- The information on past projects accumulated in Phase 1 is carried over to Phase 2.

### Collecting and summarising the information about past projects

Since the other members of your group may invest in many projects, the computer will organize the information about these past projects for you.

Instead of showing you the individual data from each of the projects that other members of your group have invested in, the information for all past projects invested in by other members of your group will be summarized in the following way. If you are Group Member C, then the computer will collect together all of these projects which have the same Attribute c value and tell you:

- (i) The number of times other members of your group (A and B) invested in a project with that value of Attribute c; and
- (ii) The proportion of projects invested in by other members of your group (A and B) with the same Attribute c value that were successful.

Specifically, Group Member C will see a screen that looks similar to the following – except with different information on success rates (other participants will see a similar screen, corresponding to their own attribute):

Figure 1: The Information Screen of Group Member C about Past Projects

## Round 3/20 (low stakes) - Please define your Investment Plan

Time left to complete this page: ⌚ 1:50

Please make an investment decision for each of the ten attribute values. You do this by selecting either "invest" or "don't invest" below each of the bars.

Attribute c value (past projects)	1	2	3	4	5	6	7	8	9	10
Percentage of past investments that were successful	0%	25%	14%	57%	9%	not enough information	25%	33%	14%	57%
Number of investments	9	12	7	7	11	0	8	12	7	7
Investment cost: €0.10	<input type="button" value="Invest"/>	<input checked="" type="button" value="Invest"/>	<input type="button" value="Invest"/>	<input checked="" type="button" value="Invest"/>	<input type="button" value="Invest"/>	<input type="button" value="Invest"/>	<input type="button" value="Invest"/>	<input type="button" value="Invest"/>	<input type="button" value="Invest"/>	<input type="button" value="Invest"/>
Success Prize: €0.34	<input type="button" value="Don't invest"/>	<input type="button" value="Don't invest"/>	<input type="button" value="Don't invest"/>	<input type="button" value="Don't invest"/>	<input type="button" value="Don't invest"/>	<input type="button" value="Don't invest"/>	<input type="button" value="Don't invest"/>	<input type="button" value="Don't invest"/>	<input type="button" value="Don't invest"/>	<input type="button" value="Don't invest"/>

Next

**Note:** In the example in Figure 1, Group Member A and B have invested in 7 projects with an Attribute c value of 4. Of these 7 projects, 57% were successful. Group Member C sees this information.

## *Timeline for decision making*

The text above has described the information that you will receive when making a decision about investing in a **single new project**. However, in the experiment, there will be 10 rounds of investment decisions in Phase 1 and 10 rounds of investment decisions in Phase 2. This makes 20 rounds of investment decisions in total. In each of these rounds of investment decision, you will be asked to state whether you would like to INVEST or NOT INVEST in **50 projects**.

In each round, the way you will do this is by telling the computer whether you would like to INVEST or NOT INVEST in projects with each of the 10 possible Attribute values (i.e. you make ten decisions and are providing the computer with an “Investment Plan” for how to act on your behalf – see the decision buttons in the bottom row of Figure 1). In each round, the computer will then see 50 randomly selected new projects, and it will follow your “Investment Plan” to decide whether to INVEST or NOT INVEST in each of these projects on your behalf. So, you make 10 decisions, and the computer uses these decisions to decide whether to INVEST or NOT INVEST in 50 projects on your behalf.

### ***Providing the computer with an “Investment Plan” for investing on your behalf***

More specifically, we will ask you to report whether you would like to INVEST or NOT INVEST in **projects with each possible Attribute value between 1 and 10**. Therefore, during each round of decisions, we will ask you to make 10 investment decisions<sup>2</sup> – one for each possible Attribute value between 1 and 10. The other members of your group will also report whether they would like to INVEST or NOT INVEST for each Attribute value between 1 and 10 for the Attribute they observe (Attribute **a** for Group Member **A**; **b** for **B**; **c** for **C**).

### ***Computer acts according to your “Investment Plan”***

In each round of decisions, after every member of the group has made their 10 decisions, each group member will face 50 randomly selected new projects. Given the choices that you have made, the computer will look at the relevant Attribute value of each of these 50 projects you face and then INVEST or NOT INVEST, according to the “Investment Plan” you gave it (i.e. if you draw a project which has an Attribute value of 6 and you said that you would like to invest in projects with an Attribute value of 6, then the computer will invest in this project on your behalf). You can change your “Investment Plan” for the next round of decisions.

### ***The “Investment Plan”: An example:***

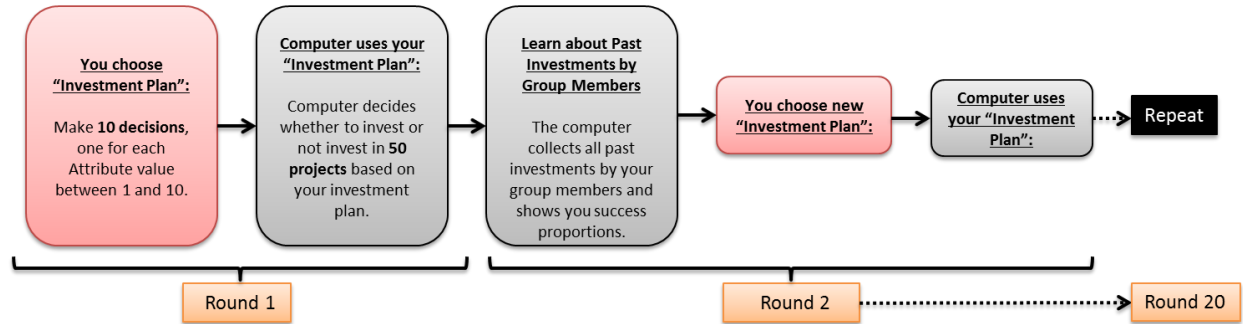
If this is a little bit complicated, it might be simpler to think about the “Investment Plan” that you give the computer in the following way. If you are Group Member **C**, and you select INVEST for an Attribute **c** value equal to 6, then you are telling the computer: “In this round, every time you see a project with an Attribute **c** value of 6, please choose INVEST on my behalf”. The computer will then go through 50 randomly chosen new projects and act on your behalf as you have instructed it.

---

<sup>2</sup> i.e. we will ask if you would like to INVEST or NOT INVEST when the Attribute value is 1; when the Attribute value is 2; when the Attribute value is 3; ...; when the Attribute value is 10.

### Updating the information you observe about past projects:

Once these 50 project decisions have been made, your database of information regarding the past success rates of projects invested in by other Group Members will be updated with their new investments, and you will be given the opportunity to revise your “Investment Plan” which will determine your investment strategy for the next 50 projects.



### How your payment will be calculated

As described above, in every round of investment decisions, you will make 10 decisions. These 10 investment decisions will provide the computer with an “Investment Plan” for how to act on your behalf when faced with the next 50 randomly drawn new projects. **In every round, one of these 50 projects will be randomly selected to be the one that affects your payment in the experiment.** For this project selected for payment, the computer will act as you have instructed it to and either INVEST or NOT INVEST. If your “Investment Plan” prescribes that the computer invest for this particular project then the contribution to your payment will depend on whether the project is *successful* or *unsuccessful*, as described above.

- Therefore, the **maximum** you can earn in Game I is:  $10 * €0.34 + 10 * €3.40 = €37.40$  if you always invest and every project you invest in is successful.
- The **minimum** you can earn in Game I is **€0** if you always invest and every project you invest in is unsuccessful.
- If you never invest in Game I, you would earn:  $10 * €0.10 + 10 * €1 = €11$ .

You will not learn about the outcomes of your own investments during the experiment. At the end of the experiment, you will be informed about the investment decisions that you made that are relevant for your payment. In each of the two phases, for the ten projects chosen to contribute to your payment, you will learn: the number of projects that you chose to invest in, the number that were successful, and how they contributed to your final payment.

**We will now proceed to Game I. Before we do, if you have any questions at this moment, please raise your hand. The experimenter will come to you.**