MICROFOUNDATIONS OF EXPLOITATION AND EXPLORATION BEHAVIOR: AN EXPERIMENTAL STUDY

REVISION-IN-PROCESS VERSION – PLEASE DO NOT CIRCULATE / CITE

Luigi Mittone (University of Trento)

Azzurra Morreale (LUT University)

Paavo Ritala (LUT University)

Abstract: Exploitation in organizations is rooted in improving existing performance, while exploration relates to the search for new alternatives and potentially higher performance. However, microfoundations of this behavior – i.e., how and why individuals exploit and explore during decision-making situations – remains largely a blind spot in the literature. We conduct an experimental laboratory study focusing on sequences of individual decision-making, including a training phase that simulates the establishment of individual-level routines, and an active phase that simulates preceding choices regarding exploitation and exploration. We financially incentivize the participants to abandon a past routine by providing higher payoffs for exploring unknown task environments. Using a sequential choice process, we observe decision-makers' choices to continue exploiting past routine, to explore (potentially earning a higher reward), or to exploit a newly established routine. We examine these choices regarding both aspiration-performance gap (the highest experienced performance and payoff), as well as recent experience (the previous performance and payoff). We find that poor performance for both of these variables increases exploration, while high performance and payoff increase exploitation of past routine as well as new routines. However, we find a difference in cumulative performance and payoffs: cumulative performance slightly increases the tendency to explore, while cumulative payoffs decrease exploration. Further, we find that decision-makers adopt markedly different patterns of exploitation and exploration behavior during sequential tasks, affecting their task performance. A pattern labelled partial exploiters provides the biggest performance and payoffs, showing how establishing and sticking to a routine quickly after an explorative search is the best alternative. Finally, we find that the availability of performance feedback improves the task performance and payoff.

1. Introduction

Exploitation and exploration are fundamental alternatives for all human decision-making. Exploitation refers to “such things as refinement, choice, production, efficiency, selection, implementation, execution” and exploration to “things captured by terms such as search, variation, risk taking, experimentation, play, flexibility, discovery, innovation” (March 1991: 71). In organizations, individual employees and managers face choices whether to exploit existing understanding and continue doing what they know works well or to explore new alternatives with hopes for improved solutions. It is, therefore, not surprising that exploitation-exploration duality has sparked a huge research program in organization sciences (Lavie et al. 2015). Some organizational tasks can be handled through previously learned routines, which refer to exploitation tasks, whereas the solutions of exploration tasks are yet unknown and thus require a search for novel alternatives and behaviors (March 1991). The main bulk of exploitation and exploration literature has remained in the organizational level (for review, see Lavie et al. 2015), loyal to the original organizational learning framing of March (1991) and Levinthal and March (1993). However, given that exploitation and exploration are fundamental processes that take place in individual cognition (Cohen et al. 2007; Laureiro-Martínez et al. 2015) and among individual decision-makers (Mom et al. 2007, 2009), it is important to know why and how individuals choose to exploit and explore.

Some recent studies aim to fill this gap in the literature by focusing on individual-level processes, which include the survey-based studies of Mom and colleagues regarding managerial tendencies to exploit or explore (2007, 2009), as well as qualitative studies of comparable managerial behavior (Stokes et al. 2015). While these studies have helped to highlight individual differences in exploration and exploitation tendencies, they still do not directly examine the behavior, its antecedents, and its possible outcomes. As a remedy, there are increasing attempts to study directly decision-making situations and behavior regarding exploration and exploitation. Billinger et al. (2014) developed an experiment where individuals’ search behavior was examined. They found that successful search promotes exploitation, while failure gradually promotes exploration. Laureiro-Martínez et al. (2015) carried out a functional magnetic resonance imaging (fMRI) experimental laboratory study, finding that exploitation and exploration activate different brain regions associated with reward-seeking, or assessment of alternatives, respectively. Using microdata from an e-commerce firm, Lee and Meyer-Doyle (2017) noted that individuals engage in more exploration when performance-based incentives are weakened. Finally, Levine et al. (2017) found in their laboratory experiment that exploitation and exploration are largely driven by past experiences, in that, exceeding aspirations drives exploitation, while trailing aspirations drives exploration.

We build on this emerging stream of literature that focuses directly on individual decision-makers' exploitation and exploration behavior. Similarly, as in the recent studies seeking to isolate exploitation-exploration choices, their antecedents, and outcomes (Billinger et al. 2014; Levine et al. 2017), we develop a laboratory experiment where decision makers face an abstracted decision-making environment, which allows testing for causality. Laboratory experiments allow for better understanding of the search behavior of individuals in complex assignments; the experimental setting enables the control and variation of external factors, such as the task's complexity, and it allows for the observation of actual human decision-making and behavior directly (Billinger et al. 2014). With our experimental design, we respond to recent calls for understanding micro-foundations of organizational behavior (Felin et al. 2012), and particularly that of bounded adaptive rationality (Puranam et al. 2015). Further, we respond to the calls to examine how i) initial experience accumulated over multiple rounds where subjects are presented with the same task, ii) exposing study participants to previously unknown explorative tasks, iii) the availability of performance feedback, as well as providing higher incentives for exploring unknown tasks, affect the decision to exploit-explore (Laureiro-Martinez et al. 2015; Levine et al. 2017). We respond to these gaps and calls by designing a laboratory experiment with 240 participants and an internal conceptual replication with two types of abstract task designs (visual and numerical). Our experimental design relies on a view of organizational behavior in which decision-makers are seen as boundedly rational and constantly updating their aspirations for feasible organizational or individual goals (Simon, 1955), referred to as *bounded adaptive rationality* by Puranam et al. (2015).

Our experimental design involves a sequential design, including a training phase where subjects play the same task over the course of eight rounds, and an active phase where subjects must decide whether to continue playing the task learnt in the training phase or explore an unknown task. This setting allows us to better understand how decision-makers face the question of how to divide attention and effort to exploiting the existing knowledge acquired during the training phase and exploring new knowledge. We assign to the participants a monetary incentive (payoff) that is computed based on the participants' capability to perform a real effort task (Brüggen and Strobel 2007), but which depends also on the decision to explore new tasks. Within this framework we examine decision-makers' choices to exploit (i.e., continue on the same task, and potentially earn the same payoff), or to explore, i.e., switch to a new task, and potentially earn a higher payoff but with a higher uncertainty about their performance. This design corresponds to the nature of exploitative and explorative learning at the individual level (March 1991; Cohen et al. 2017). Further, the existence of a training phase in our experiment provided the decision-makers a set of routines as a basis (similarly as decision-makers in organizations; Nelson and Winter 1982), thus providing an additional opportunity, which we call a exploitation of past routine, as opposed to exploitation of new routines that are developed during the experiment. Importantly, the use of a "real-effort task" (Brüggen and Strobel 2007) helps in reproducing the development of routines as a function of repetition and task success (Weiss and Ilgen 1985 p. 59). Following Laureiro-Martinez (2014, p. 1112), when implemented, routines "require considerably less time and cognitive resources than exploring new and unknown solutions". Applications of real-effort tasks are experiments where tasks are repeated and subjects typically use their experience acquired in the first periods to gradually improve their performance over the course of the periods (Benndorf et al. 2019). Therefore, real-effort tasks are perfectly suitable for our experimental setting.

Our study contributes to the growing literature of individual-level exploitation and exploration behavior via a unique research design that allows to confirm and elaborate the results of other studies, as well as provides more insights on relevant contextual factors and behavioral patterns of exploitation and exploration. First, our results confirm the findings of both Billinger et al. (2014) and Levine et al. (2017), in that, we also find that poor performance (reflected against the best previous performance) promotes exploration. These performances reflect the aspiration-performance gap that is known to affect managerial choice (March 1988). We complement these findings by also examining the direct effect of the absolute performance in the previous round, which suggests that recent poor performance facilitates exploration and recent good performance exploitation (see also van Rijnsoever et al. 2012). Second, we test the role of cumulative experience for both task performance and payoff. This reflects the slack search as discussed in behavioral theory (Cyert and March 1963), and which was recently tested in an experiment by Levine et al. (2017). Our results provide interesting findings that complement these views. Similar to Levine et al. (2017), we find that cumulative high payoffs (i.e., payoff slack) reduce exploration tendency. On the other hand, we find partial evidence that high performance slack increases exploration. Here, it might be that the decision-makers deviate in their slack search behavior when they have “money in the bank” (payoff slack) and a high level of confidence in their past performance (performance slack), but a motivation to search for higher payoff. Third, unlike previous studies, we test how the availability of performance feedback (i.e. reduced uncertainty about performance) affects task performance. As known in the previous organizational literature, such performance feedback helps to improve individual performance (Larson 1984). Our results confirm that decision-makers reach higher performance and payoffs when they have performance feedback available, which is likely to improve their choices over whether to exploit or explore. Finally, and differently from previous studies, our research design allows us to construct behavioral patterns of exploitation-exploration choices over a sequence of actions. We find that the decision makers adopt five different strategies, which we label as *consistent exploiters, retreaters, partial exploiters, rational explorers,* and *unstable explorers.* We find that the highest performance and payoff are related to the pattern of partial exploiters, which are decision-makers who change just once or twice, and then establish a new routine that continues to harness a higher payoff related to an early-on explorative choice.

2. Exploration-exploitation choice at the individual level: bounded adaptive rationality view

The seminal literature on exploitation and exploration is grounded in the organizational level and was initially concerned about organizational learning (March 1991; Levinthal and March 1993), even if it was inspired by the early ideas of Simon (1955), regarding adaptive aspirations and individuals' exploration choices. Laureiro-Martinez et al. (2010), among others, noted that the majority of the studies focusing on exploration and exploitation begin with March's (1991) seminal paper and target their analysis on higher levels than the individual. March's (1991) work looks at the trade-off of exploration and exploitation in a social setting within organizations and the two different elements, namely, mutual learning within the organization and the competition for primacy among organizations. The underlying notion of exploration and exploitation is a turnover process that includes an assortment of individuals with different cognitive characteristics, where some are natural explorers and some have a tendency toward exploitation (see also Laureiro-Martinez and Brusoni, 2018). Correspondingly, the majority of organizational literature focuses on the organizations' benefits and challenges as it comes to exploitation and exploration (e.g., Hoang and Rothaermel 2010; Lavie et al. 2015), as well as its simultaneous pursuits to achieve both at the same time (Smith and Tushman 2005).

Increasingly, organizational literature is starting to identify that there are differences among the individual decision-makers in terms of their abilities in exploitation and exploration (Mom, Van den Bosch and Volberda 2007; Mom, Van den Bosch and Volberda 2009). The merits of individual-level investigation have been shown in recent experimental designs that have demonstrated how the past performance matters in exploitation-exploration choices (Billinger et al. 2014; Levine et al. 2017), and in studies that examine the micro level patterns of behaviors under different performance incentives in an organization (Lee and Meyer-Doyle 2017). Recent studies using neuroimaging approaches further support these findings, showing clear differences between exploitation and exploration behavior (Laureiro-Martinez et al. 2010, 2015). However, as Puranam et al. (2015: 352) point out, the "conditions under which people switch between exploration and exploitation are not yet fully understood".

We follow the original definition of March (1991: 71), where exploitation refers to "such things as refinement, choice, production, efficiency, selection, implementation, execution" and exploration to "things captured by terms such as search, variation, risk taking, experimentation, play, flexibility, discovery, innovation". Building on this seminal definition, we conceptualize exploitation-exploration choice in the individual level. In terms of exploitation, the key issue is that the task is known by the decision maker based on previous experience, and thus a process of learning and refinement can take place (Ericsson and Lehmann 1996). Regarding exploration, the opposite is true: a decision-maker decides to abandon the current task in order to find a new, potentially more profitable one (March, 1991; Laureiro-Martinez et al., 2015; Smith & Tushman, 2015). This disengagement gives the opportunity for "the pursuit of new knowledge, of things that might come to be known" (Levinthal & March, 1993: 105). We view exploitation and exploration as a sequence of choices by the decision maker, where both types of behavior serve particular performance goals – learning-by-repetition on known tasks vs. experimentation on new and uncertain, but potentially more profitable tasks. This is a dynamic process: every occurrence of exploration has later a potential to become exploitation, after the same task is repeated. On the other hand, after several repetitions of (even successful) exploitation, decision makers might choose to explore new alternatives. While organizational scholars have suggested the tendency to overexploit due to preference of existing competences (Levinthal & March, 1993), recent experimental research has also shown the tendency to overexplore, where performance improvements in local task have been cut short by too early disengagement (Billinger et al., 2014).

In examining individual-level behavior in exploitation and exploration, we ground it to two interrelated streams of literature: the Herbert Simon's view of decision-maker as a boundedly rational agent with adaptive aspirations (Simon 1955)[1], and the literature that has followed under the label of behavioral theory and problemistic search (Cyert and March 1963; Gavetti et al., 2012). Importantly, the canonical trade-off between exploitation and exploration (March 1991) itself is linked to the notion of a boundedly rational decision maker and to dynamic sequential decision-making processes. Together, these refer to the process of exploring better alternatives by reacting to performance feedback and in settling to exploit satisfactory performance (Cyert & March, 1963; Simon, 1996)[2]. This boundedly rational search model resonates well with the micro-behavioral view in recent experimental studies of exploitation and exploration (Laurero-Martinez et al., 2015).

In our experimental approach, we follow the *bounded adaptive rationality* model as described in Puranam et al. (2015: 341-346). First, the *task environment* involves the objective reality of a decision maker where a set of possible actions which map into different performance outcomes. The decision-maker is assumed to adopt actions that lead an acceptable performance, or in other words, reach a certain aspiration level (Cyert and March 1963; Posen et al., 2018). Furthermore, while the performance itself is objective, the decision maker has a representation of it based on the beliefs over a set of feasible actions and related performance outcomes. Second, a *choice process*, given the representation, leads the decision maker to choose an action which pursues the aspired performance outcomes. This choice process is often incomplete, given the limited availability of information about performance determinants or the abilities of the agent to perform. Furthermore, the process does not always lead to maximization of absolute performance, but rather in "satisficing": i.e. calibrating a choice related to the decision-maker's aspiration level (Simon, 1955; Cyert & March, 1963). Third, *feedback* is generated from the consequences of decision maker's action in the task environment. The feedback relates to performance, and its availability and quality varies. Fourth, the feedback triggers a *transformation* of the decision maker's representation of the task environment based on performance feedback over a task (Puranam et al., 2015). This has consequences for the choice process itself. For instance, the problemistic search literature predicts that when a decision maker perceives poor performance based of the feedback, this triggers a search for alternative (Greve, 1998; March 1988). On the other hand, when performance feedback is good, the decision maker is drawn to repeat such behavior that lead to that performance. Overall the bounded adaptive rationality view to exploitation-exploration choice views it as a dynamic process where decision-maker chooses compares the different alternatives and adapts the aspiration level accordingly. This choice is adapted by the feedback the decision maker has accumulated at the time of the decision, including both recent as well as previous experience.

3. Experimental approach

Our experimental approach builds on the above discussed model of bounded adaptive rationality, and implements several important elements that help us to model organizational reality of exploitation-exploration choice. Importantly, we recognize that individuals *learn* by experience of task environments and from the performance feedback received. For instance, Ericsson and Lehmann (1996, p. 290) discovered in their laboratory study that task performance clearly enhances after training as individuals' skills improve (see also Posner et al. 1997). Referring to such findings, Helfat and Peteraf (2015) conclude that such path dependence in the development of cognitive capabilities can explain heterogeneity in the potential as well as actual performance of mental activities of decision-makers (see also Laureiro-Martinez & Brusoni, 2018).

To accommodate learning by experience to our experimental design, we develop a model with specific tasks, defined as a set of actions and choices aimed at realizing a specific target (Puranam et al., 2015). The degree of fulfillment of the target determines the level of achievement of the objective function of the decision makers. In this model, exploitation takes place as the deliberate repetition of a particular task. The repetition leads to learning, and ultimately can lead to *individual-level routines*, with sufficiently reliable performance in a given task environment (Laureiro-Martinez, 2014: Oehler et al., 2019).[3] Exploration is then the deliberate search of a new task. In our experimental setting the task is repeated many times (rounds), in each rounds the rules of the task do not change, while it changes (only to a small extent) the specific target. The small change of the value of the target from one round to the next mimics an unstable environment that is often considered one of the challenges for developing efficient routines, in this sense we follow the idea of "appropriateness" in shaping a routine (March, 1992). Round after round by experiencing different combinations of actions and choices the decision maker eventually can learn an individual best strategy of executing the task and can fully *learn* to exploit the potential performance available in the task. To decide to explore a new task is equivalent to look for a new set of available choices and actions, keeping at the same time unchanged the objective function of the decision maker (i.e. the best approximate of an always changing target).

We assume that the decision maker follows a dual decision process: 1) choosing and implementing the right actions to extract the performance potential from the available task (in which a routinization takes place over a set of rounds), and then 2) decide whether to continue exploiting the same task or to explore new ones. Importantly, we define exploitation the decision to continue on the same task and not the repetition of the same actions and choices within each round. Similarly, we define exploration as the decision to abandon a task for a new one and not the decision to try different strategies for performing the same task. This allows the model learning and routinization, where the decision maker is able to learn a new task-performing procedure if the newly explored task is exploited in the preceding rounds. It is also useful to note that we expect the decision makers become more efficient in executing the same task across repetitions (exploiting it) because they are learning a better task-performing strategy (see e.g. Ericsson and Lehman, 1996).

In our theoretical and experimental setting, the degree of realization of the objective function and the consequent reward for the decision maker depends on the level of performance achieved in carrying out the task and on the monetary reinforcement (potential payoff) associated with the degree of approximation to the target associated with the task. In our experimental setting the reinforcement is calibrated to make it relatively less convenient to repeat the same task for a large number of rounds, assuming that the individual ability to perform tasks does not vary with changing tasks. This experimental feature is intended to motivate exploration and allows to mimic the core of the exploitation-exploration dilemma (March, 1991), where exploitation generates more certain and costless benefits but can generate inertia, while exploration offer potential higher benefits but expose to higher risks (Greve, 2007). Exploiting a well-known task is safer than exploring a new one so the reasons why the decision maker might decide to explore must be driven by a mixture of favorable expectations with respect to the set of actions offered by the new task and with respect to a higher reward. In our setting the decision maker is not interested in improving his performance as such but is aimed at obtaining a higher level of achievement of his objective function. The decision maker objective function depends from the performance but is also influenced by the monetary reinforcement associated to the task and coincide with his monetary compensation (personal reward). In this sense we follow the “induced value Theory” by Vernon Smith (1976) which is considered as a standard technique in Experimental Economics.

To ensure consistency in our dual individual decision-making setting we assume that the decision makers use the same ontology to take their choices in both the two decisional levels. In this sense, we adopt a behavioral dynamic perspective within which the decision maker initially become “expert” of an exogenously assigned task and then can decide to stick on the familiar task (exploiting it) or to explore a new task, chosen from a set of new alternative ones. Consistently with the model of bounded adaptive rationality (Puranam et al. 2015) we also assume that the decision maker is driven by the same sequential and interrelated process of comparison between aspiration level and realization of goals. We integrate this baseline model with the behavioral pattern delineated in *aspiration adaptation theory* (Selten, 1998) that allows for three decisional patterns: exploiting, exploring and retreating (to past exploitation).

The adoption of this theoretical framework allows to distinguish between two types of exploitation: *exploitation of past routines* (EPR) and *exploitation of new routines* (ENR). Importantly, both types of exploitation build on learning via repetition, but with a distinction that we enforce in the experimental setting regarding routines already possessed vs. deliberate choices to build new routines. EPR corresponds to the exploitation of a task that has been learnt via repeated experiences on that task, leading to a routinized performance (Laureiro-Martinez 2014; Oehler et al. 2019). ENR implies that the decision maker is exploiting a new routine – including deliberate repetition of a task that has been explored as an alternative to a past routine. *Exploration*, then, is a deliberate choice of a new task that deviates from a routine that is currently being built (ENR) or a past routine (EPR).

The whole behavioral process (Figure 1) starts with an initial stage where the decision makers learn a routine, through a repetition of the same task, then they must decide if they want to get stuck on the same routine (EPR) or if they prefer to switch to a new task (exploration). If the decision makers decided to explore a new task, they can thenceforth decide to reuse the new task, starting to learn a new routine (ENR) or to return to the initial task, or to explore another new task. Thus, we also see that the decision to return to a past routine involves a choice of "retreating" (Selten 1998), where the decision maker deviates from a novel choice back to a known behavior already learned across a set of past experiences.

INSERT FIGURE 1 HERE

As Figure 1 shows, a choice taken at round t + 1 (where t is the number of rounds during the routinization of a given task x) is conditioned by the feedbacks generated during the accumulation phase of the routinized behavior (from round 1 to t). At round t + 1 the decision maker must decide to exploit the past routine or to explore. From round t + 2 onward the if the decision maker had chosen to explore in round t + 1 he can decide to retreat, exploiting the routinized past task x, or to exploit the recent task which explored in the previous round t + 1, or can decide to continue to explore[4].

In the following sections, we shortly build arguments regarding the role of recent experience, the gap between current and highest previous performance (aspiration-performance gap), the role of cumulative experienced performance (i.e., performance slack), as well as availability of performance feedback.

3.1 Recent experience

Recent literature of exploration, exploitation, and search behavior examines the role of performance feedback signals based on previous experiences (e.g., Van Rijnsoever et al. 2012; Billinger et al. 2014; Levine et al. 2017). According to Van Rijnsoever et al. (2012), the assumption is that the decision to explore is primarily influenced by the most recent experience, while older experiences progressively lose their impact on the decision makers' choice process. This explanation holds even if the working memory of decision maker affects this process (Laureiro-Martinez et al. 2019). In particular, if the outcome of the most recent choice is perceived as a failure, it is more likely that the decision-maker will not perform that choice again, and is more prone in exploring a different choice. Conversely, if the outcome of the most recent choice is perceived a success, a repetition of the same choice is expected (Van Rijnsoever et al. 2012; Lave and March 1993; Sitkin and Pablo 1992). We expect this to hold particularly strongly with exploitation of new routines, where the decision-maker has recently explored an alternative successfully as well as set a new aspiration level, leading to exploitation of that same choice.

3.2 Past highest performance (aspiration-performance gap)

While the most recent experience clearly matters, behavioral stream of literature suggests a role for longer-term aspiration levels (Lant and Montgomery 1987; Greve 1998; Cyert and March 1963; Levitt and March 1988; March, 1988). This means that the decision maker revises the *aspiration level* in consideration of all the feedback received in the past in comparison to recent performance (Billinger et al. 2014). This intuition is supported by prospect theory by Kahneman and Tversky (1979) and suggests that decision makers evaluate the results obtained from previous choices on the basis of a subjectively fixed reference point. Following Billinger et al. (2014) and Levine et al. (2017), the highest-performing experience, among the past experienced ones, is used as a reference point, thus allowing us to consider aspiration of performance enhancement (Billinger et al. 2014; MacLeod and Pingle, 2005). According to Greve (1998; see also March 1988), if current performance is below the aspiration levels, then the likelihood to explore increases. When individuals perceive the current performance in comparison to the reference point, that is, if the current performance is lower than the reference point, it is perceived as failure (Bromiley 1991; March 1988). A failure to meet aspirations leads to risk-seeking and a more explorative behavior, and in turn exceeding aspiration levels leads to risk aversion and a more exploitative behavior (Billinger et al. 2014; Levinthal and March 1993). In our setting, we further assume that while a negative aspiration-performance gap promotes exploration, it decreases exploitation of both past routine as well as new ones. In contrast, experiencing similar or higher performance than the previous reference point is likely to promote both types of exploitation.

3.3 Cumulative performance and experience (slack)

In our experimental design, we examine a sequential process of choices, where performance as well as experience choices accumulate. This corresponds (in the micro-level) to the exploitative and explorative learning approach of March (1991) as well as the adaptive process of boundedly rational search (Simon, 1955). This cumulative process of trial and error also coincides with the "reinforcement learning theory" (Sutton and Barto 1998; Cohen et al. 2007). Indeed, the reinforcement learning theory represents a crossroads among different scientific disciplines. Sutton and Barto (1998, p. 4) state, "*One of the challenges that arise in reinforcement learning, and not in other kinds of learning, is the trade-off between exploration and exploitation. To obtain a lot of reward, a reinforcement learning agent must prefer actions that it has tried in the past and found to be effective in producing reward. But to discover such actions, it has to try actions that it has not selected before... ...The agent must try a variety of actions and progressively favor those that appear to be best.*"

The mechanism described by Sutton and Barto (1998) supports the intuition previously presented in Figure 1. More precisely, it implies the existence of a trade-off between the cumulative knowledge of the consequences produced by past choices and the decision to exploit past routine, explore new alternatives, or exploit newly discovered alternatives. In fact, this kind of decision maker discussed must explore in order to find the actions that "*appear to be best.*" This strategy should consider, in some way, like in standard searching models (Stigler, 1961), the tradeoff between the expected costs and the expected advantages deriving from exploring a new choice. One could then assume that a series of repetitions of the same choice should trigger the decision maker to assume the outcomes (feedbacks), which were obtained as a benchmark for defining her aspiration level.

In fact, and not surprisingly, the role played by a sequence of positive feedback, and by the degree of salience of past experiences, has been highlighted in literature. The foundation for these discussions is that of *slack search* (Cyert and March 1963), which was recently conceptualized in an experimental setting as "performance slack" (Levine et al. 2017). This concept relies on a consolidated debate in organization theory and recalls, quite closely, the reinforcement learning theory. Performance slack in organization literature concerns situations where external circumstances, that have an impact on the organization, tend to reproduce over time, so reducing the willingness of organizations to explore (Levinthal and March 1993). Alternatively, the tendency to exploit can also explain, via positive feedback, the results for perfecting a specific type of organizational behavior: as decision makers get positive feedback from their organizational environment on a choice that they have recently pursued, they are likely to improve further on those aspects, resulting in better performance (Sydow et al. 2009). When a stabilizing performance is witnessed, and experience of good performance cumulated, the decision makers could be expected to lock into those behavioral patterns. In summary, we can argue that long sequences of positive performance should trigger the decision maker to continue exploiting the same choices. In contrast, when cumulative negative performance accumulates, other behavioral choices such as exploration or retreating to a past routine become more likely.

3.4 Availability of performance feedback

Decision-makers, directly or indirectly, witness performance feedback all the time based on their actions. However, the availability of useful performance feedback varies a lot. We know from the organizational literature that, in many situations, the individual decision-making process takes place in a condition of incomplete, or even total absence of, information about the immediate consequences produced by the actions performed (e.g., Ritala et al. 2016). Some decisions are literally “made in the dark”, where decision makers only know much later, if even then, the performance implications of their decisions (e.g., Weick 1993).

Organizational and economic experiments have widely analyzed the role of decision feedback. For instance, Diehl and Sterman (1995) found that as the delay between cause and effect grows, the decision-making performance suffers. Among the most important ways to manipulate information in experiments involve instances where the participants receive different information settings to create an asymmetric information environment, for example in simple bargaining contexts (e.g., Roth and Malouf 1979) or in studying market equilibrium dynamics (e.g., seminal paper by Smith 1962). This evidence suggests that different levels of the accuracy and amount of the information provided to the participants often interfere with their capability to choose the best options and strategies. It follows that, in our setting, we could assume that such a role could be played by information and, more precisely, that the more information available to decision makers, the better is their choice process.

4. Research design

This section describes the implemented experimental design, including the tasks involved in the experiment, the study participants, as well as the experimental procedures used to produce data on the individuals' choices related to exploitation and exploration in their decision-making performance.

4.1 Experimental design

Our experiment relies on a computer game, which involves a "real effort task" (Brüggen and Strobel 2007) which allows us to examine the actual decision-maker's skills and their accumulation during the experiment. In order to test the adaptive aspirations among decision-makers, the game was designed as sequential where participants altogether played 18 rounds. The game is divided into two main phases: the training phase (8 rounds) and the active phase (10 rounds). Aligned with our theoretical logic, the training phase simulates the accumulation of an individual-level routine on a specific task environment, while the active phase simulates sequences of an exploitation and exploration choice, where adaptive aspirations play a role. During the training phase, all of the study participants are presented with the same task, which they perform for 8 rounds. Conversely, in the active phase, subjects can switch during each round to a different and not known task from among three alternative tasks that are made available or, instead, stay with the same task they played in the training phase. Specifically, during the training phase, subjects are randomly assigned to a "*building*" task (where they learned a technique for building a visual puzzle or a number) or to a "*memory*" task (where they learned a technique for memorizing a puzzle or a number). This procedure allows the subjects to acquire distinct routines in two different task environments. Independently from the specific task, the general goal function consisted in approximating each and every round of the experiment, in the best possible way, as a “target” assigned to the participants. In the context of decision making, the target simulates any type of goal that the decision maker receives performance feedback on and, relatedly, chooses to exploit or explore. We used two frames for representing the goal function:

In the *visual frame*, the participants were requested to select or construct an abstract “target figure.” Similar to the study conducted by Mittone and Papi (2017), all the figures we use in this experiment are grids with cells painted red and beige (see Table 1). In the paper, we denote each cell as a *pixel* and a set of four adjacent cells as a *block*. The *Numeric frame's* design corresponds closely to the visual frame, but only in a different context. Also, here, the participants were requested to select or construct a “target number.” In total, our experiment relies on 4 tasks: one learned in the training phase and the other 3 unknown tasks that, in addition to the learned one, could be chosen by subjects during the active phase. In order to not affect the choices, we used neutral – rather than descriptive – terminology to present the tasks. Table 1[5] elaborates a detailed description of the designed tasks used in the present experimental study.

INSERT TABLE 1 HERE

Notably, in any round, people are presented with a different figure or number. That is, during the training phase, they become confident with the task (which is always the same), and the outcome is predictable and not completely certain. This gradual learning process allows us to capture the definition of exploitation by March (1991, p. 81) as “*the refinement and extension of existing competences, technologies, and paradigms. Its returns are positive, proximate, and predictable.”* Given our setting, in the ninth round (i.e., in the active phase), subjects faced the dilemma of whether to choose (i.e. *exploit*) the same task learned in the training phase (whose outcome is familiar but still uncertain) or to *explore* a new and unknown one (whose outcome is largely uncertain). From the tenth round on, if participants decided to explore a new task in round nine, they faced three choices: 1) exploit the past routine by retreating to the same task learned in the training phase, where the task is known with certainty; 2) exploit a new routine, i.e. the task just explored recently in the active phase; or 3) explore another task.

Two alternatives to informing about task performance: Feedback and no-feedback

In order to test for the availability of information on performance feedback, two different versions of the game were created, which differ in the system of giving feedback. In the *feedback* version, participants received numerical feedback at every round regarding their performances both during the training and the active phase. Indeed, consistent with previous research (Levine et al. 2017, van Rijnsoever et al. 2012) and with several organizational situations (Gupta, Smith and Shalley, 2006) in our design, and during the active phase, a decision-maker chooses and performs the task, receives performance feedback, and chooses once more. Specifically, participants received feedback about their performance, but not about the payoffs[1], reproducing an environment where managers often operate. We replicate this version under both the numerical and visual scenarios. In the *no feedback* version, the participants were not informed about their performance in any round of the game (although they are informed of the payoff scheme in the beginning). Thus, we can observe how the absence of performance feedback affects exploration-exploitation behavior. Similar to the study where participants receive feedback, we replicate this version under the numerical and visual scenarios.

Performance computation

In all versions, the performance was computed in terms of the distance of the result from the target figure (the target number), according to the following formula:

$$Fitness(\%) = 100 - \left(100 * \frac{d_i(c,s)}{T}\right)\%$$

Specifically, in the visual frame, $d_i(c,s)$ is the number of mistakes made by subject *i* for choice problem *c* in the scenario *s*=*visual*. Following Mittone and Papi (2017), in our design a mistake is counted at any time a pixel of the chosen (or constructed) figure is different from the corresponding pixel of the target figure. In the numerical frame, $d_i(c,s)$ is the difference, in absolute value, between the number

[1] They are informed about the potential payoff scheme at the beginning of the game, but not after each round.

selected (or constructed) by subject *i* for choice problem *c* in the scenario *s=numerical* and the target number. In the visual frame, T is equal to $d_{max}(c, s)$, i.e., the maximum number of mistakes, a subject can commit f or choice problem *c* in the scenario *s=visual*. In the numerical frame, T is equal to the target number. It is important to highlight that in the building (destroy) task embodied in the visual frame, wherein people do not use the available blocks and leave the figure as it is, i.e., partitioned into blank (red) spots, the performance is higher than zero by definition. That is, some pixels are already correctly positioned compared to corresponding pixels in the target figure, so that the initial performance is higher than zero. Therefore, in order to have the numerical frame comparable, we assumed that the initial performance is the same as it is in the visual game by placing a given number in the provided formula.

Payoff computation

Participants were monetarily incentivized according to the standard paradigm used in Experimental Economics (Smith 1976). To pay the participants, we computed their performances in accordance with the degree of correct approximation to the targets, and then in the end, the performance was transformed into *experimental currency units* (ECU) using the conversion in Table 2 (the participants were fully informed about it).

INSERT TABLE 2 HERE

However, during the active phase, if a participant decided to switch from the well-known task and explore a new unknown one, s(he) obtained a potentially higher payoff (see Table 3).

INSERT TABLE 3 HERE

The use of a higher performance-ECU ratio is justified by the idea that to change a well-known task with a new one should be in some way incentivized to compensate the costs implied by the learning of a new routine. Furthermore, in this way, we could mimic the *potential* (but not certain) advantages implied by the decision to explore new tasks (March, 1991). In other words, to decide to explore a new task exposes the participant to an uncertain result, which depends upon the (unknown) difficulty of the new task.

Risk Elicitation

After having played the main experiment (that is, the afore-mentioned training and active phases), subjects were also asked to play the BRET (Bomb Risk Elicitation Task) in order to elicit their risk preferences (Crosetto and Filippin, 2013). In this task, subjects are presented with 64 boxes and informed that, among the boxes, there is one box containing a bomb, while the remaining 63 boxes contain 1 ECU each. Subjects are asked to collect the number of boxes they want. It is known that if the box containing the bomb is collected, it explodes, nullifying the earnings for this part of the experiment. Boxes are then opened. If the box with the bomb has not been selected, the subject's earnings increase with the number of collected boxes. Conversely, if the box containing the bomb has been collected, the subject's earnings are null. The higher the number of collected boxes, the higher the propensity for risk taking.

4.2 Participants and experiment procedures

We conducted our experiment in an experimental laboratory of an European-based university. Participants were recruited via a customized software implemented at the laboratory. The experiments were programmed and run by using the "o-tree" software (Chen et al. 2016; Holzmeister et al. 2016). In total, 240 subjects took part in the experiment, of which 132 were female and 108 were male. The average age of the participants was 21.80 (s.d. = 2.36), and most of the participants were students of economics (54.17%) and the others were students of law (21.25%), engineering (7.5%), the social sciences (4.58%), the humanities (3.33%), psychology (1.67%)or mathematics and other natural sciences (7.50%).

In the experiment, we ran the *Numerical-feedback (NF)* study in four sessions, the *Visual-Feedback (VF)* study in four sessions, the *Numerical-No feedback (NNF)* study in two sessions, and the *Visual-No feedback (VNF)* study in two sessions. Each subject participated in one study only. Specifically, out of 240 people participants, 81individuals took part in the *NF* study, 81 individuals in the *VF* study, 39 individuals in the *NNF* study, and finally 39 individuals in the *VNF* study.

Upon arrival to the laboratory, subjects were randomly assigned to a computer, where they read the instructions related to the first part of the experiment (the "training phase"). They were told that the experiment was composed of two main phases, that is the "training phase" and the "active phase."

Specifically, it was common knowledge that they would participate in the second part of the experiment (the "active phase"), but none of them knew in advance the purpose of this second part. After completing the training phase, subjects received the instructions related to the active phase. Finally, subjects were asked to play the Bomb Risk Elicitation Task (BRET; Crosetto and Filippin 2013) and to complete a short demographic questionnaire.

In addition to 3 Euros as a show-up fee, subjects received an amount of experimental currency units (ECU), in accordance with the choices made both in the first and second parts of the experiment (training and active phase) as well as in the BRET. Specifically, participants were informed that at the end of the experiment two rounds would be randomly extracted (one from the training phase and one from the active phase), and they would be payed according to the results that they had obtained in those specific rounds. At the end of the experiment, the experimental currency units were then converted to Euros (100 ECU = 1 Euros). The average payment was €11.07 (show-up fee included), and the experiment lasted approximately 40 minutes.

5. Results

In this section, we report the findings from our experimental study. We first present the analysis of data concerning both the performance and response time evolution while participants played the training phase. Then, we present regression analyses of the determinants of explorative-exploitative behavior. Finally, we investigate the behavioral patterns and performance implication.

5.1 Routines in the training phase

Following the arguments presented in sections 1 and 3, we argue that, the existence of a training phase in our experiment provided the decision-makers to establish a routine. Specifically, if a routine is implemented during the training phase, subjects should complete the task in less time and should be more efficient, that is, they should improve their performance over time (Laureiro-Martinez, 2014). Accordingly, we checked this issue computing both the average performance round by round during the training phase and the average response time needed by the participants to performance their task during the training phase. In line with the fact that participants, on average, routinize, we observe that the average performance of the participants increases from round 1 to round 6 and then stabilizes in rounds 7 and 8 (see Figure 2). Moreover the average response time needed to take decisions diminishes during the training phase (see Figure 3). Interestingly, we notice that, compared to subjects who played in the no-feedback version, participants that received feedback at every round regarding their performances, not only, on average, completed the assigned task faster, but they also performed better in any round of the training phase.

INSERT FIGURE 2 HERE

INSERT FIGURE 3 HERE

5.2 Determinants of exploitation and exploration behavior

Before introducing our regression analyses, we define, in detail, how both independent and dependent variables were operationalized (for summary, see Table 4).

INSERT TABLE 4 HERE

Dependent variables

Given our experimental setting, we define *exploration* as the behavior that implies switching to a task which is different from the one performed in the previous round and is not a task involved in the training phase. On the other hand, choosing the same task as in the previous round, or choosing the same task learned in the training phase, is classified as *exploitation*. More precisely, we distinguish between two types of exploitation—*exploitation of past routine* (every time that participants, during the active phase, decide to remain or retreat to the training task) and *exploitation of new routine* (every time that participants, during the active phase, decide to repeat a task different from the one learned in the training phase). Both *exploration* and *exploitation of past routine* are operationalized as dummy variables (see Table 3), allowing us to compare those with alternative behavior in two sets of analyses.

Independent variables

The independent variables used in our regression analysis follow the general theoretical framework discussed into the previous sections. Here, we assume that the decision maker follows a satisficing strategy, adapting the aspiration level on the basis of the results obtained from the past experiences, including performance and payoffs. Thus, to test the effect of the most recent experience, we measure the performance as well as the payoff experienced in the previous round (*last performance/payoff experienced*).

We also included two variables inspired by the most recent experimental literature (Billinger 2014; Levine et al. 2017), which attributes an important role in explaining the decision to explore the past best result obtained by the decision maker as a proxy of *Aspiration-performance/payoff gap* and the average cumulative result in terms of performance (*Performance slack*) and payoff (*Payoff slack*). The intuition behind the introduction of these variables is that the participants should adapt their strategy according to a process of aspiration adaptation, which is not only influenced by the results observed in the previous round but, instead, also influenced by their overall experience across the whole experiment. Our econometric model also includes a set of dummy variables that allows us to isolate any effect induced, respectively, by the different frames (variable *Game*), the presence/absence of feedback performance (variable *Feedback*), and the kind of training task that the participants played from round 1 to round 8 (variable *Training*).

Finally, we add the variable *BRET* (that allows us to control for the impact induced by subjects' risk preferences), *Age* (in years), a dummy variable for the gender, a dummy variable for those studying economics (*Economics)*, and *Number of experiments* (in terms of participation in previous experiments).Regarding risk, the literature on decision-making behavior has long identified that individuals have different risk-attitudes (see e.g., Diamond and Stiglitz 1974), which may affect explorative or exploitative behavior. Regarding gender, behavioral economics literature shows that female subjects behave differently from male subjects; for instance, women are, on average, more financially risk averse than men (Charness and Gneezy, 2012). Also, exploration-exploitation literature has found evidence for different behavior between males and females (Mehlhorn et al. 2015). The variables *Economics*, *Age*, and

Number of experiments have been included to control for possible idiosyncratic effects triggered by demographic differences among the participants.

Regression results

Tables 6 and 7 display the results of logit regressions (including individual random effects) on the entire sample of data, with consideration of the observations during the active phase. Specifically, Model 1 and Model 7 examine how the independent variables "*last performance experienced*", "*negative performance-gap,*" and "*performance slack*" influence both exploration and exploitation of past routine. It is apparent that, in the case of higher recent performance (i.e. performance experienced in the previous round), subjects are less likely to explore and thus more likely to exploit. We also notice that higher recent performance slightly reduces exploitation of past routine. Taken together, these findings suggest that recent high performance increases the tendency to exploit a task different from the learning task (i.e., exploitation of new routine), while poor performance increases exploration. For the role of the *aspiration-performance gap*, we find that negative gap increases exploration and decreases exploitation of past routines. In terms of effect sizes, a one standard deviation increase in negative *Aspiration-performance gap* increased the likelihood of exploring by 764%, while the same increase reduced the likelihood of exploiting a past routine by 98%. These findings offer strong support for March's proposition on the role of negative-aspiration-performance gap in pushing exploration and boosting exploitation. Further, we find that cumulative performance (*performance slack*) does not affect either exploration nor exploitation.

Models 2 and 8 incorporate the control variables, including risk-attitudes (BRET), feedback (dummy), numerical task (dummy), and the memory task in the training phase instead of building phase. It is apparent that the aforementioned variables have the same role as in models 1 and 7, except for the variables "*last performance experienced"* (that is no longer significant in model 8), and for *performance slack*, which increases exploration by 2% (Model 2) and slightly reduces exploitation of past routine by 5% (model 8). We also find that the numeric task slightly increases exploration by 33%, while it does not affect the likelihood to exploit a past routine. Also, being trained on a memory task strongly reduces exploration while increasing exploitation of past routine at the same time. Specifically, the likelihood of exploring and exploiting a past routine decreases and increases by 55% and 689.90 respectively. An explanation of this result can be found in the fact that during the training phase, subjects, on average, performed better when trained on a memory task (average performance = 98,52) rather than on a building task (average performance = 84,78)[2]. That is, when subjects become very confident with a learned routine, they are more likely to continue or retreat to it. Finally, the coefficients of the variables *Gender, Feedback,* and *BRET* are not statistically significant. These considerations hold when we also control for *Economics*, *Age,* and *Number of experiments* (models 3 and 9).

Models 4, 5, 6, 10, 11, and 12 investigate on the role of payoff (instead of performance) on our dependent variables. Specifically, Models 4 and 10 examine how *last payoff experienced, negative payoff-gap,* and *payoff slack* impact both exploration and exploitation of past routine. It is apparent that the results are consistent with what we already found when such independent variables were computed in the domain of performance (Models 1 and 7). In particular, we have again evidence of the fact that if current payoff is below the aspiration levels, then the likelihood to explore increases, while the likelihood to exploit decreases. In fact, a failure to meet aspirations is associated with an average increase of 234% (Model 6) in the likelihood of exploring and a 66% change in the likelihood of exploiting a past routine (Model 12). Likewise we confirm the sizeable effect of the specific task learned during the first 8 rounds: being trained on a memory task reduces exploration by 42% (model 6) while increasing exploitation of past routine by 208,56.

The only exception is represented by effect of the cumulative payoff (*payoff slack*), which reduces exploration and has no effect on exploitation of past routine (models 4 and 10). This result is consistent with what Levinthal and March (1993) name “success trap,” that is, “money in bank” reduces further search and increases exploitation. Adding control variables in models 5 & 6 and 11 & 12 does not change the effect of these three afore-mentioned variables, except for the variable *payoff slack*, which not only reduces

[2] The average performance in the memory task (98.53) is statistically and significantly higher than the average performance in the building task (84.78) (p value = 0.000 computed with a Mann–Whitney U test).

exploration but also exploitation of past routine. Thus we found that *payoff slack* especially induces exploitation of new routine. However, the effect is relatively small: a one standard deviation increase in payoff slack reduces the likelihood of exploring and exploiting a past routine by 1%.

Importantly, we confirm the findings already discussed in the domain of performance that regard the additional variables.

INSERT TABLE 5 HERE

INSERT TABLE 6 HERE

5.3 Behavioral patterns in exploitation and exploration

In order to identify the main behavioral patterns within the decision sequences of the players, we first ran a cluster analysis to check if the switching strategies used by the participants were distributed in a random way or if some kind of natural clusters existed that would allow us to distinguish between fundamental styles. We used a variable that measures the number of changes (switches from one task to another) completed during the active phase to run the cluster analysis. To this variable, we then added a second variable that measures which type of task the participants chose during the active phase. As discussed in the experimental design section, we used four tasks: *Building*, *Memory*, *Combined*, and *Destroy*. Combining the kind of task chosen together with the number of changes done allows us to obtain clusters. We ran the cluster analysis by restricting the data-set only to rounds 9 to 12 because including all the rounds of the active phase generated an excessive number of clusters and the majority of the participants (more than 55%), after round 13, stacked on the same choice strategy. We identified three main clusters. Table 7 reports the results from the three clusters in regard to the number of switches done by the participants:

INSERT TABLE 7 HERE

Overall, it is noticeable that the three clusters separate the more repetitive styles of playing from the more dynamic ones, going from cluster one (prevalently repetitive) to cluster three (highly dynamic). Cluster 2 can be seen as a sort of "mixed" group, having a mix of repetitions of the same task as well as switching among tasks. To go more deeply into the understanding of the individual behavioral patterns, we decided to make a further classification of the strategies used by the participants. In this way, we also resolved some of the overlaps among the clusters shown in Table 7. Following the theoretical premises already described in the theoretical section, we can classify our participants into six categories as follows.

Consistent exploiters are participants who never switch from the task learned during the training phase. The assumption here is that the participants belonging to this category feel satisfied, in terms of aspiration level, with the results obtained from the training task, and, therefore, they did not aspire to improve their outcome (payoff) by trying new tasks. This category includes 1 subject who switched in the last round and one subject who, in round 14, switched once, only immediately come back to the training task. *Rational explorers* are participants who apply a logic very close to what Simon would have defined as global rationality. In fact, at the beginning of the active phase (specifically between rounds 9 and 12), they tested all three new tasks (remember that the new tasks were unknown) and then locked in the one that they believed was the best among the four experienced (including the task learned during the training phase) for all the remaining rounds of the experiment. This decision implies that they fixed their aspiration level after the first four trials (having, in this way, "experienced" all four of the available tasks) and never revised it anymore. Alternatively, we can suppose that they applied a standard goal maximization rationality by choosing the best task after testing them all, i.e., the task that allowed them to maximize their expected payoff. *Unstable explorers* are participants who switch across the different tasks more than 4 times and across the whole duration of the active phase. They can be imagined as decision makers who are continuously revising their aspiration level. During this repeated process of switching, they would sometimes also retreat to the training task. *Partial exploiters* explored only one or two tasks at the very beginning of the active phase (i.e., at round 9 and 11) and then stacked on that task until the end of the experiment. We assume that this behavior is triggered by a low aspiration level that induced them to minimize the effort needed to switch from one task to another. This category also includes 2 subjects who, just in the last round, decided to break the new routine (choosing a different task), and 3 subjects who, just in one round (between 10 and 11), decided to come back to the routine behavior. *Retreaters* only once explored a new task at the very beginning of the active phase (i.e., between round 9 and 11) and then decided to go back to the training task for the entire length of the game. *Others* are all the participants who cannot be included in the five main typologies (as described above), and whose behavior involves very diverse variety of exploitation and exploration decisions. In Table 8, we report the descriptive statistics regarding the key features of the game design and participants, including training task type, gender, as well as whether they received accurate performance feedback or not.

INSERT TABLE 8 HERE

Looking at the effect produced by the training phase, one can easily notice that the training for the memory task has a strong effect on the percentage of *Retreaters*. In fact, 87% of the whole group of *Retreaters* received training using the memory task. Table 8 also reports a similar result in regard to the type of *Consistent exploiter* (76% of the whole group received training using the *memory* task). This finding suggests that some participants perceived the memory task as "easy," or at least easy enough to discourage exploration. We could, therefore, think that the nature of the task learned during the training phase, or *ceteris paribus*, the degree of perceived advantageousness of a task learned and routinized, is a key variable in influencing the attitude to explore. Moreover, it is worth underlining that this last consideration is tied to our theoretical assumption of a boundedly rational agent. Following this line of interpretation, it is not surprising that the *Rational explorers* type splits exactly in half between those who have been trained with the *memory* task and those who were trained with the perceived more difficult *building* task. In fact, to be "trapped" into the training task requires a strategic approach deeply rooted in a limited capability to elaborate on alternative scenarios and aspirations.

Alternatively, we can see some irregularities between different frames (visual or numerical) and the behavioral patterns, which do not follow a particular intuition. In terms of gender, the distribution among males and females across types is quite homogenous. The only exception to this general consideration is represented by the types *Others* and *Rational explorers*. The gender composition of these two types is reversed with a majority of females in the type *Others* and a majority of males in the type *Rational explorers.* Finally, it is possible to observe that there were some minor differences in feedback and no feedback scenarios among the groups.

Table 9 compares the average performances and payoffs (considering observations from round 9 on), in addition to risk-preferences attitudes (BRET) among different behavioral patterns.

INSERT TABLE 9 HERE

Considering all variables, the six groups do not have equal variance according to the *Bartlett* test. Thus, average performance, payoff, and risk attitudes are significantly different among the six behavioral types. Looking at the post-hoc analysis based on pair-wise comparison between the behavioral groups (last column of Table 9), it is possible to observe that all categories, except for the *Rational explorers,* present a value of performance that is significantly higher than the performance realized by the participants belonging to the type "*Others*" (that present the lowest value of performance). In terms of payoff, it is possible to observe that *Consistent exploiters*, *Retreaters*, and *Others* gained a significantly lower payoff than the one obtained by the participants belonging to the other behavioral groups. It is interesting to notice that these three typologies are not homogenous in regard to playing strategies. More precisely, we can say that the fundamentally conservative strategy played by both the *Retreaters* and the *Consistent exploiters* produces the same (poor) result obtained by the "chaotic strategy" implemented by the participants belonging to the *Others* typology. This result recalls the intuition related to a "*success trap*" (Levinthal and March 1993) that, at the end, confirms the negative side of over-exploitation. In other words, one can state that the absence of a sharp strategy of choice, as well as a tendency to stick on the most "comfortable" and routinized task, produces, in the long run, a progressive worsening of the economic performance (of course, under the assumption that some kind of more profitable task does exist).

In contrast, the *Partial exploiters* category reached the highest payoff, which shows an average payoff that is significantly higher than the average payoff of all the other typologies. It is worth remembering that participants were financially motived to choose—in the active phase—a task different from the task played during the training phase. Our results show that exploring one or two new tasks at the very beginning of the active phase, and then continuously exploiting it (*Partial exploiters*), seems to be the best strategy in terms of payoff. In fact, the *Partial exploiters'* strategy maximizes two advantages: the first one is represented by the payoff multiplier "activated" by the decision to abandon the training task, and the second one is the opportunity to learn the new task from the very beginning of the active phase. In this way, the *Partial exploiters* can quickly reach the same level of performance already reached in the training task but using a task that has a potentially higher payoff.

Furthermore, it is intuitive that the most risk averse typology is also one of the least explorative one: the *Retreaters* show the lowest value for BRET and also are the poorest overall group in payoff. In addition, this result directly points to many literature evidences that show how the decision to exploit (in this case retreating) is typically linked to a lower propensity to assume the risks. Conversely, the high-risk attitude reported by the "purest" category of exploiters, meaning the *Consistent exploiters*, seems to go counter to the previous intuition.

More generally, in order to analyze the effect of several factors on both performance (Models 13 and 14, and 15) and payoff (Models 16, 17, and 18) achieved by participants during the active phase, we decided to conduct six random effects generalized least squares (GLS) regression models, introducing random effects to check for repeated decisions (Table 10). The independent variables include the six behavioral groups and the control variables previously considered, that is, BRET, a dummy variable for the specific game played (numeric against visual), a dummy variable for the presence of performance feedback, a dummy variable for the gender, a dummy variable for the training task (memory against building), *Economics*, *Age,* and *Number of experiments*. The group "Others", Female, visual game, and the absence of performance feedback are our baseline categories.

INSERT TABLE 10 HERE

As Model 13 in Table 10 indicates, the other behavioral patterns outperform the "others" group, with the exception of *Rational explorers*, which does not show a significant difference. Models 14 and 15, including control variables, show that the *Feedback* coefficient is positively, and significantly, correlated with performance, suggesting that the providing performance feedback, at any round, leads to higher levels of performance. The coefficients of the variables *BRET, Gender,* and *Game* are not statistically significant; therefore, they do not have a role in the explanation of our dependent variable. Alternatively, as shown in Model 16, the groups *Unstable explorers, Rational explorers,* and *Partial exploiters* achieved a higher payoff than the group *Others*. On the contrary, *Retreaters* got a significantly lower payoff, while there is no difference between *Others* and *Consistent exploiters* in terms of payoff. Model 17 and 18, with additional variables, substantially align with the above results and shows that getting performance feedback leads to higher payoffs In addition, playing the numeric game leads to significantly higher payoffs. Finally, *Gender BRET, Economics, Age, and Number of experiments* are not statistically significant.

6. Discussion and implications

Our study focused on examining the classic organizational processes of exploitation and exploration (March 1991) as individual-level phenomena. We join the nascent, but growing stream of studies, that utilize experimental designs to examine decision-makers' choices as they relate to explore and exploit (e.g., Billinger et al. 2014; Laureiro-Martinez et al. 2015, 2019; Levine et al. 2017). Our experimental design complements this stream of studies by adding several important specifications that allow us to examine closely the decision-making behavior and, by doing so, to contribute to organizational literature regarding micro-foundations of exploitation and exploration.

In particular, our experimental design combines the intuition of bounded adaptive rationality models (Simon, 1955; Puranam et al., 2015) with Selten's (1998) aspiration adaptation theory and assumes that different feedbacks, received from previous experiences, affect the decision to deliberately exploit a recent choice, retreat to a well-known past routine, or to explore a new choice. The underlying assumption of this model is that the decision maker carries on a continuously adapting process of comparison between aspirations and performance feedback. Examining and explaining this variety of alternative choices is important, as shown in a recent experimental study by Laureiro-Martinez et al. (2019): they found that longer term “working memory” of managers, regarding their past choices, allows them to do better exploitation-exploration decisions. Thus, we argue that it is important to understand the full variety of past behavior as well as the accumulated experience in order to explain different forms of exploitation and exploration. To do so, we developed an abstract experimental laboratory setting that allows us to study the microfoundations (Felin et al. 2012) of exploitation-exploration choices, focusing on the actual behavior rather than the outcomes or indirect ques of this behavior. This involves tasks that the experiment participants go through in *sequential rounds*, receiving feedback from their immediate performance, and having the possibilities to continue exploiting the known task (switch to new tasks, or later retreat to familiar tasks). This allows us to examine closely diverse *behavioral patterns*, where participants adopt different approaches to whether and when they exploit and explore. Using the data regarding the sequential choices, we are able to examine the antecedents of exploitation and exploration choices, both as a result of very recent performance and as performance that took place earlier on.

Our experiment focused on actual exploitative and explorative learning (March 1991) by adopting a real effort task (Brüggen & Strobel 2007), in that the participants are trained to use their own skills, and the uncertainty they face relates to the extent of their skills in further the exploitation or exploration of a new task (rather than a randomized lottery or similar, as often used in experiments). To this end, we included a *training phase* in our experiment, which creates a base level of experience, similar to a typical organizational setting where individuals face choices whether to continue in their routinized behavior or ignite a search for alternative solutions (e.g., Gavetti and Levinthal 2000). Past experience can greatly influence the performance of the later choices (Ericsson and Lehmann 1996; Helfat and Peteraf 2015); thus, testing the role of experience in an experimental setting is an important addition. This addition is in contrast to several related experimental studies that involve either abstract lotteries (Laureiro-Martinez et al. 2015) or tasks where the decisions to explore refer to being exposed to a random performance in the performance landscape (e.g. Levine et al. 2017). Our design, in turn, follows a setting where learning takes place during the task (similar to Billinger et al. 2014) and allows participants to test the choices, as well as task performance, in a setting that resonate well with the organizational reality where the learning and skills by the decision-maker are crucial in determining performance.

Similarly, as in typical experimental designs, we financially motivated the participants to perform to the best of their abilities. However, in our design, we specifically provided higher potential payoffs to incentive abandoning of routines, following the conceptualization of exploration as a higher risk, higher reward activity (March 1991). This design allowed us to test to which extent the participants behave in a "rational" way, and also which types of patterns of behavior generate the best performance and payoffs.

Based on previous literature, we expected that a positive recent experience increases the likelihood of exploitation. As expected, the results demonstrated that recent high performance and payoff increases exploiting that same choice, while, at the same time, likelihood to explore new tasks, or to exploit a past routine (task was learnt in the training phase) are reduced. We also expected that, similar to Levine et al. (2017), that a negative aspiration-performance gap increases exploration. We found that a low current performance, as well as payoff, compared to a previous maximum level predicts an explorative choice for the next task. At the same time, in these cases continuing to exploit the recent "sub-optimal" task is reduced, as is the tendency to retreat to exploiting a past routine. We also examined the role of cumulative performance and payoff in exploitation-exploration choice, expecting it to induce exploitative behavior. This discussion parallels to that of slack search (Cyert and March 1963), where the performance in the past creates an accumulated "slack" of resources, affecting behavior. On the one hand, we found a difference in cumulative performance and payoffs: performance slack increases tendency to explore, while, on the other hand, payoff slack increases exploitation (and particularly only short-term exploitation). This demonstrates that, on the individual level, slack search involves different processes that are contingent on whether it is experienced good performance (performance slack) or "money in the bank" (payoff slack).

Finally, relying on a data-driven categorization of different behavioral patterns of exploitation and exploration, we examined the performance and payoffs of individual decision-makers. We found that a pattern labelled *partial exploiters* was the most successful behavioral pattern. In this pattern, the decision-makers explore once or twice and, after this, quickly settle on exploiting a high level of performance. This finding supports a view of boundedly rational behavior (Simon, 1955, 1997), where aspiration is dynamically adjusted and, when a satisfactory alternative is found, an exploitative behavior ensues. Relatedly, we involved a manipulation of the accuracy of *performance feedback*. We found that decision-making is improved with more information about performance of choices, where the availability of accurate and immediate performance feedback improves the average performance and the payoffs.

We replicated the laboratory design in two abstract decision-making settings, the visual and the numerical, as well as using four types of tasks, in order to examine the extent to which the behavioral processes of exploitation and exploration are generalizable across problem domains. While controlling for the types of tasks across the analyses, we also conducted post-hoc tests examining how differences in the initial training task (which establishes a routinized behavior) affects performance later. We found strong evidence of "over-exploitation" (Billinger et al. 2014), or a "success trap" (Levinthal and March 1993; Rhee and Kim 2014), type of behavior. When receiving an easier task (here: the memory task, see also Ericsson and Lehmann 1996) to establish a routine on, the decision-makers tended to continue exploiting this task and did not explore potentially higher-reward alternatives, resulting in a poorer payoff.

All in all, our study contributes to better understanding of the micro-foundations of a key organizational phenomenon of exploitation and exploration. The focus on micro-level phenomena is not new but has its origins in early management theory (Felin et al. 2012). Cyert and March (1992), as well as March and Simon (1958), in their behavioral theory of the firm, explained the organizational heterogeneity with microfoundational issues. Barnard (1968, p. 139) argued that "the individual is always the basic strategic factor of organization." Thus, microfoundations should be a focal determinant for explicating not only heterogeneity of firm performance (Powell et al. 2011) but also the emergence of different capabilities and processes (Felin et al. 2012).

References

Barnett WP, Hansen, MT (1996) The red queen in organizational evolution. *Strategic Management Journal* 17:139–157.

Benndorf V, Rau, HA, Sölch C (2019) Minimizing learning in repeated real-effort tasks. *Journal of Behavioral and Experimental Finance*, 22: 239-248.

Billinger S, Stieglitz N, Schumacher TR (2014) Search on rugged landscapes: An experimental study. *Organization Science*, *25*(1), 93–108.

Bromiley P (1991) Testing a causal model of corporate risk taking and performance. *Academy of Management Journal*. 34(1):37–59.

Brüggen A, Strobel M (2007) Real effort versus chosen effort in experiments. *Economics Letters*, 96(2), 232-236.

Charness G, Gneezy U (2012) Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, *83*(1):50 –58.

Chen DL, Schonger M, Wickens C (2016) oTree - An open-source platform for laboratory, online and field experiments. *Journal of Behavioral and Experimental Finance,* 9:88–97.

Cohen JD, McClure SM, Yu AJ (2007) Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):933–942.

Crosetto P, Filippin A (2013) The "bomb" risk elicitation task. *Journal of Risk and Uncertainty,* 47(1):31–65.

Cyert RM, March JG (1963) *A Behavioral Theory of the Firm* (Prentice-Hall, Englewood Cliffs, NJ).

Diamond PA, Stiglitz JE (1974) Increases in risk and in risk aversion. *Journal of Economic Theory* 8(3):337-360.

Diehl E, Sterman JD (1995) Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes*, 62(2):198-215.

Ericsson KA, Lehmann AC (1996) Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, 47(1):273-305.

Felin T, Foss NJ, Heimeriks, KH, Madsen TL (2012) Microfoundations of routines and capabilities: Individuals, processes, and structure. *Journal of Management Studies*, *49*(8):1351-1374.

Gavetti G, Levinthal D (2000) Looking forward and looking backward: Cognitive and experiential search. *Administrative Science Quarterly*, 45(1): 113–137.

Gavetti G, Greve HR, Levinthal DA, Ocasio W (2012) The behavioral theory of the firm: Assessment and prospects. *Academy of Management Annals*, 6(1): 1-40.

Giddens A (1991) *Modernity and self-identity: Self and society in the late modern age.* Stanford University Press.

Greve HR (1998) Performance, aspirations, and risky organizational change. *Administrative Science Quarterly* 43:58–86.

Greve HR (2007) Exploration and exploitation in product innovation. *Industrial and Corporate Change*, 16(5): 945-975.

Gupta AK, Smith KG, Shalley CE (2006) The interplay between exploration and exploitation. *Academy of Management Journal* 49(4):693–706

Helfat CE, Peteraf MA (2015) Managerial cognitive capabilities and the microfoundations of dynamic capabilities. *Strategic Management Journal*, *36*(6):831–850.

Hoang HA, Rothaermel FT (2010) Leveraging internal and external experience: exploration, exploitation, and R&D project performance. *Strategic Management Journal*, 31(7):734 –758.

Holzmeister, Felix, Pfurtscheller, Armin (2016) oTree: The "bomb" risk elicitation task. *Journal of Behavioral and Experimental Finance*, Elsevier, vol. 10(C):105–108.

Kahneman D, Tversky A (1979) Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–291.

Lant TK, Montgomery DB (1987) Learning from strategic success and failure. *Journal of Business Research* 15:503–517.

Laureiro-Martinez D (2014) Cognitive control capabilities, routinization propensity, and decision-making performance. *Organization Science*, 25(4): 1111-1133.

Laureiro-Martínez D, Brusoni S, Canessa N, Zollo M (2015) Understanding the exploration–exploitation dilemma: An fMRI study of attention control and decision-making performance. *Strategic Management Journal*, *36*(3):319 –338.

Laureiro-Martínez D, Brusoni S (2018). Cognitive flexibility and adaptive decision-making: Evidence from a laboratory study of expert decision makers. *Strategic Management Journal*, 39(4): 1031-1058.

Laureiro-Martinez D, Brusoni S, Tata A, Zollo M (2019) The manager's notepad: working memory, exploration, and performance. Journal of Management Studies. In press. DOI: 10.1111/joms.12528.

Larson Jr JR (1984) The performance feedback process: A preliminary model. *Organizational Behavior and Human Performance*, 33(1):42–76.

Lave CA, March JG (1993) An Introduction to Models in the Social Sciences, 2nd ed. University Press of America, Inc., London.

Lavie, D, Stettner, U, Tushman, ML (2010). Exploration and exploitation within and across organizations. *Academy of Management Annals*, 4(1): 109-155.

Lee S, Meyer-Doyle P (2017) How performance incentives shape individual exploration and exploitation: Evidence from microdata. *Organization Science*, *28*(1): 19–38.

Levinthal DA, March JG (1993) The myopia of learning. *Strategic Management Journal*, 14, Special Issue, 95–112.

Levine SS, Reypens C, Riedl C (2017) Should I stay or should I go? The cognition of exploration and exploitation. The Cognition of Exploration and Exploitation. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2816541.

Levitt B, March JG (1988) Organizational learning. *Annual Review of Sociology* 14:319–340

MacLeod WB, Pingle M (2005) Aspiration uncertainty: Its impact on decision performance and process. *Journal of Economic Behavior & Organization* 56(4):617–629.

March JG (1988) Variable risk preferences and adaptive aspirations. *Journal of Economic Behavior & Organization*, 9(1):5–24.

March JG (1991) Exploration and exploitation in organizational learning. *Organization Science*, *2*(1):71–87.

March JG (1992) *A Primer on Decision-Making: How Decisions Happen*. NY: the Free Press

Mehlhorn K, Newell BR, Todd PM, Lee MD, Morgan K, Braithwaite VA, ... & Gonzalez C (2015) Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, *2*(3):191.

Mittone L, Papi M (2017) Does inducing choice procedures make individuals better off? An experimental study. *Theory and Decision*, *83*(1):37 –59.

Mom TJ, Van Den Bosch FA, Volberda HW (2009) Understanding variation in managers' ambidexterity: Investigating direct and interaction effects of formal structural and personal coordination mechanisms. *Organization Science*, *20*(4):812 –828.

Mom TJ, Van Den Bosch FA, Volberda HW (2007) Investigating managers' exploration and exploitation activities: The influence of top-down, bottom-up, and horizontal knowledge inflows. *Journal of Management Studies*, *44*(6):910 –931.

Oehler PJ, Stumpf-Wollersheim J, Welpe IM (2019) Never change a winning routine? How performance feedback affects routine change. *Industrial and Corporate Change*, 28(5): 1101-1124.

Posen HE, Keil T, Kim S, & Meissner FD (2018) Renewing research on problemistic search—A review and research agenda. *Academy of Management Annals*, 12(1): 208-251.

Powell TC, Lovallo D, Fox CR (2011) Behavioral strategy. *Strategic Management Journal*, 32(13):1369–1386.

Puranam, P, Stieglitz, N, Osman, M, Pillutla, MM (2015). Modelling bounded rationality in organizations: Progress and prospects. *Academy of Management Annals*, 9(1), 337-392.

Ritala P, Heiman B, Hurmelinna-Laukkanen P (2016) The need for speed—unfamiliar problems, capability rigidity, and ad hoc processes in organizations. *Industrial and Corporate Change*, *25*(5):757–777.

Roth AE, Malouf MW (1979) Game-theoretic models and the role of information in bargaining. *Psychological Review*, 86(6):574.

Selten R (1998) Aspiration adaptation theory. *Journal of Mathematical Psychology*, *42*(2-3):191–214.

Simon HA (1955) A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1):99 – 118.

Simon, H. A. (1996). *The Sciences of the Artificial* (3rd ed.). Cambridge, MA: MIT Press.

Simon HA (1997) *An Empirically Based Microeconomics*, Cambridge, UK: Cambridge University Press.

Smith VL (1962) An experimental study of competitive market behavior. *Journal of Political Economy*, *70*(2):111–137.

Smith VL, (1976) Experimental Economics: Induced Value Theory, *The American Economic Review*, Papers and Proceedings of the Eighty-eighth Annual Meeting of the American Economic Association, 66(2):274-279

Smith WK, Tushman ML (2005) Managing strategic contradictions: A top management model for managing innovation streams. *Organization Science*, 16(5):522–536.

Stigler GJ (1961) The economics of information. *Journal of Political Economy*. 69(3):213–225.

Sitkin SB, Pablo AL (1992) Reconceptualizing the determinants of risk behavior. *Academy of Management Review* 17:9–38.

Stokes P, Moore N, Moss D, Mathews M, Smith SM, Liu Y (2015) The micro-dynamics of intraorganizational and individual behavior and their role in organizational ambidexterity boundaries. *Human Resource Management* 54(S1):s63–s86.

Sutton RS, Barto AG (1998) *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA.

Sydow J, Schreyögg G, Koch J (2009) Organizational path dependence: Opening the black box. *Academy of Management Review* 34(4):689 –709.

Van Rijnsoever FJ, Meeus MT, Donders ART (2012) The effects of economic status and recent experience on innovative behavior under environmental variability: An experimental approach. *Research Policy* 41(5):833–847.

Weick KE (1993) The collapse of sensemaking in organizations: the Mann Gulch disaster, *Administrative Science Quarterly* 38:628–652.

Weiss HM, Ilgen DR (1985) Routinized behavior in organizations. *Journal of Behavioral Economics* 14(1): 57-67.

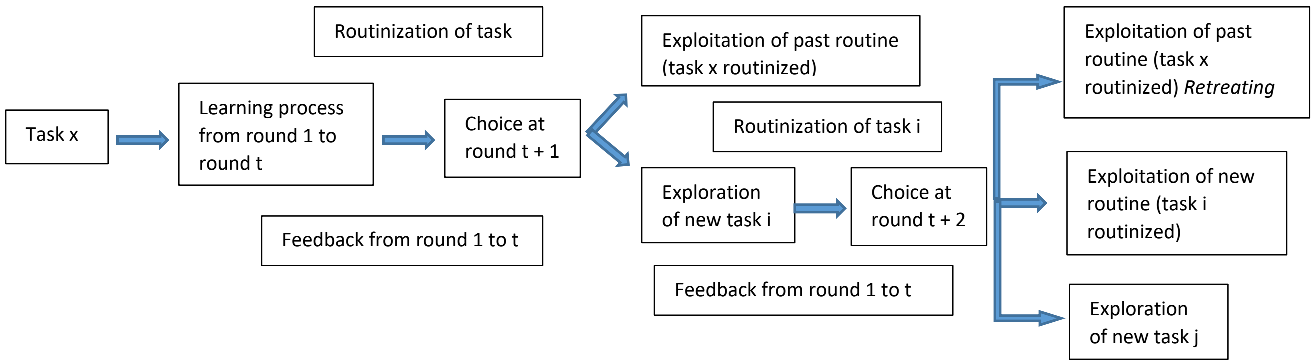


Figure 1. Temporal framework of exploitation and exploration behavior

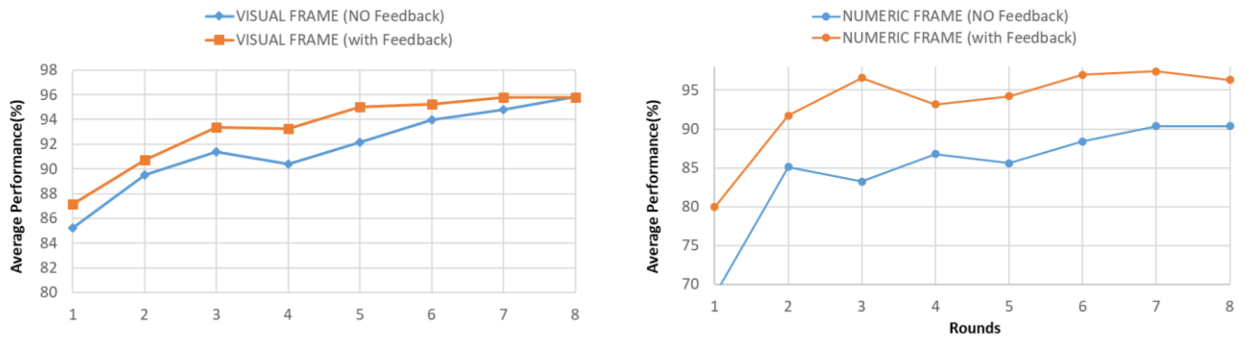


Figure 2 Average Performance during the training phase (rounds 1-8. On the left, visual frame, on the right numeric frame

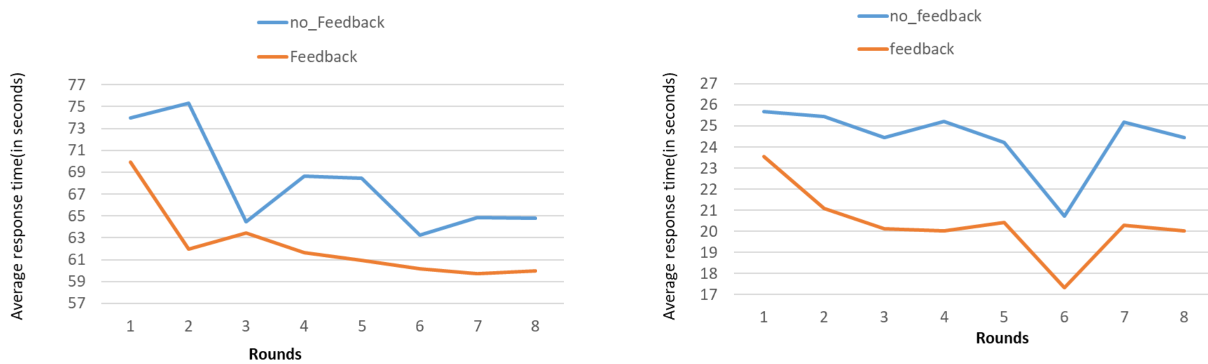
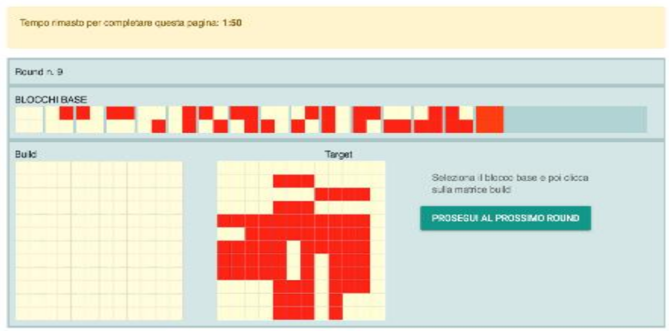
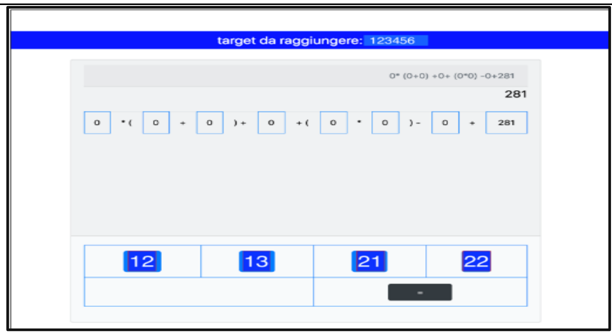
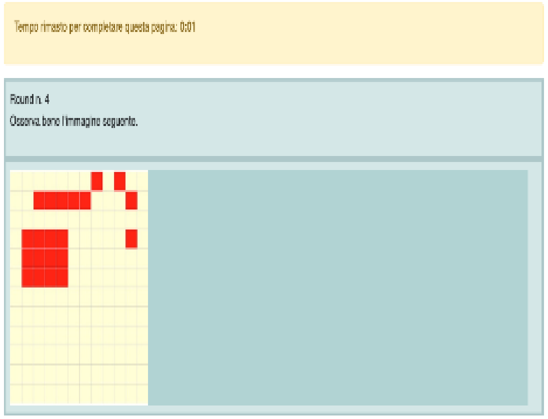


Figure 3 Average Time response (in seconds) during the training phase (rounds 1-8). On the left, visual frame, on the right numeric frame

Table 1. Task description in the visual and numerical games

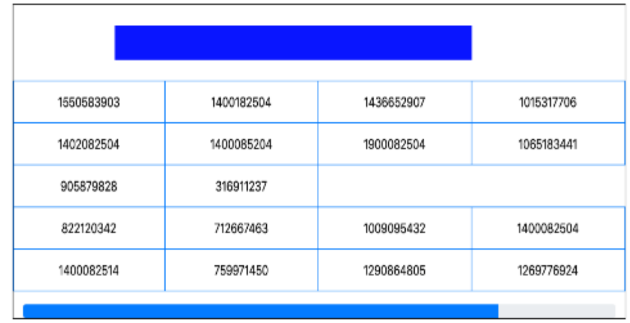
<table>
<tr><th>Task description</th><th>Visual Game</th><th>Numerical Game</th></tr>
<tr>
<td><u>Building task</u> – participants must construct, to the best of their ability, a target figure or number (visual or numeric) by using a set of blocks in the visual frame (and numbers in the numerical frame), which is shown for the entire duration of each round on their computer screens.</td>
<td>


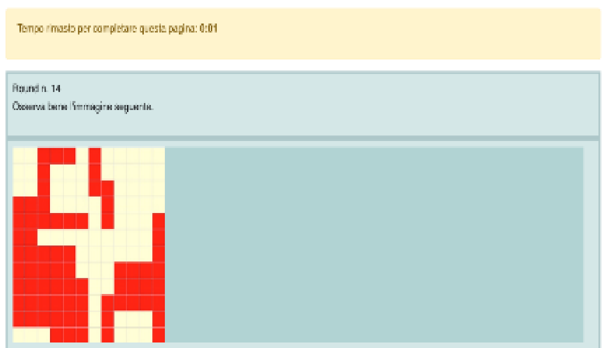
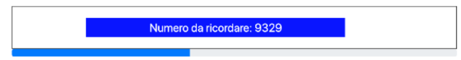
Subjects are shown several blocks, a target figure (on the right side), and one figure (on the left side) partitioned into blank spots of the same dimensions as the blocks. Starting from the figure on the left side, subjects are asked to construct a figure by placing the blocks in the slots within 120 seconds. Before the time expires, participants are allowed to allocate and replace any block to any slot.</td>
<td>


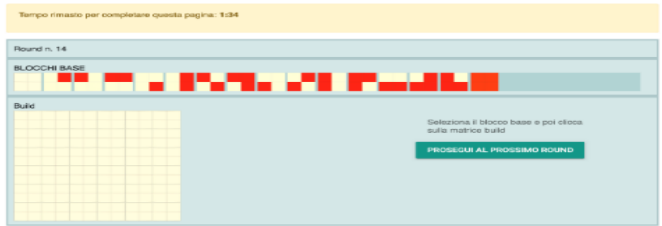
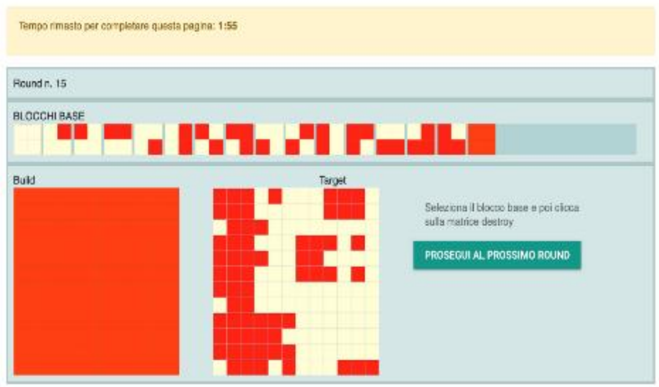
Subjects are shown a target number (on the top of the figure) and four numbers (along the bottom of the figure). Subjects are asked to construct a number using a pre-determined formula by placing the available numbers into pre-fixed empty cells within 60 seconds. Before the time expires, participants are allowed to allocate and replace any number to any cell.</td>
</tr>
<tr>
<td><u>Memory task</u> – in this task participants are shown a target figure or number for a pre-determined time (20 seconds in the visual frame and 5 seconds in the numerical frame).</td>
<td>


</td>
<td>


</td>
</tr>
</table>

<table>
<tr>
<td>Then, the target is hidden and participants are presented with a set of 18 figures (including the target one) or 18 numbers (including the target one), and they must choose the one that they believe is the target one. Participants have to make such a choice within 120 seconds in the visual frame and 60 seconds in the numerical one.</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td><u>Combined task</u> is a combination of the memory task and the building task. Participants are presented with a target figure or number for a given time (20 seconds in the visual frame and 5 seconds in the numerical one).

Then the target is hidden, and they are requested to build it in the same way as in the building task, with the difference that this time the figure or number target is not shown any more during the building phase.</td>
<td>
</td>
<td>
</td>
</tr>
</table>

<table>
<tr>
<td>Participants have to construct the figure or the number within 120 seconds in the visual frame and 60 seconds in the numerical one.</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td><u>Destroy task</u> is the mirror-like task of the building task. The only difference is that subjects are shown a figure partitioned into red spots (rather than beige) in the visual frame, and a filled formula in the numerical task.</td>
<td>

Subjects are shown several blocks, a target figure (on the right side) and a figure partitioned into red spots with the same dimensions as the blocks (on the left side). Starting from the figure on the left side, subjects are asked to construct a figure by placing the blocks in the slots within 120 seconds. Before the time expires, participants are allowed to allocate and replace any block to any slot.</td>
<td>

Subjects are shown a target number (on the top of the figure) and four numbers (along the bottom of the figure). Subjects are asked to construct a number using a pre-determined formula by placing the available numbers into pre-fixed filled cells within 60 seconds. Before the time expires, participants are allowed to allocate and replace any number to any cell.</td>
</tr>
</table>

Table 2. Conversion scheme performance-Experimental Currency units (ECU)

Performance(%)	ECU
100% - 95%	400
94% - 80%	200
79% - 65%	70
64% - 50%	20
49% - 0%	0

Table 3. Conversion scheme performance- Experimental Currency units (ECU) if the task chosen is different from the task learnt in the training phase

Performance (%)	ECU
100% - 95%	600
94% - 80%	300
79% - 65%	100
64% - 50%	30
49% - 0%	0

Table 4. Variables and measurement

Definition	**Measure explanation**
Dependent variables	
Exploration	Binary variable [0,1]. 1 if the participant changed tasks from the previous round but has not retreated to the task learned in the training task, 0 otherwise.
Exploitation of past routine	Binary variable [0,1]. 1 if the participant remained in or returned to the same task learned in the training task, 0 otherwise.
Independent variables	
Performance slack	Continuous variable: Average of the accumulated performances from the beginning of the game to the current round.
Last performance experienced	Continuous variable[0.100]: performance in the previous round
Aspiration-performance gap	Binary variable [0,1]. Computed as the difference between the performance of the current round and the maximum performance reached across all the past rounds. It takes a value of 1 if the difference is negative, 0 otherwise.
Payoff slack	Continuous variable: Average of the accumulated payoffs from the beginning of the game to the current round.
Last payoff experienced	Continuous variable: payoff in the previous round
Aspiration- payoff gap	Binary variable [0,1]. Computed as the difference between the payoff of the current round and the maximum payoff reached across all the past rounds. It takes a value of 1 if the difference is negative, 0 otherwise.
BRET (risk-measure attitude)	Number of boxes collected in the "Bomb Risk Elicitation Task (BRET). Higher values of BRET are related to a higher degree of risk taking
Game	Binary variable [0=visual, 1=numeric], whether the experimental frame is visual or numeric
Gender	Binary variable[0=female,1=male]
Feedback	Binary variable[0 = no, 1 = yes], whether participants received performance-feedback or not
Training task	Binary variable[0=build, 1=memory], whether the task played in the previous 8 rounds was a building task or a memory one

Age	Continuous variable (in years)
Economics	Binary variable [0, 1]. It takes a value of 1 if participants have a background in Economics, 0 otherwise
Number of experiments	Continuous variable (how many experiments the subject participated to previously)

Table 5a. Determinants of exploration (performance)

Dependent variable: exploration	**Model 1**	**Model 2**	**Model 3**
Last performance experienced	-0.113*** (0.000)	-0.013*** (0.000)	-0.0135*** (0.000)
Aspiration-performance gap (Negative gap=1)	2.156*** (0.000)	2.21*** (0.000)	2.21*** (0.000)
Performance Slack	0.0009 (0.917)	0.021* (0.022)	0.0185* (0.041)
Feedback (Yes=1)		-0.172 (0.308)	-0.1699 (0.315)
Game (Numeric=1)		0.285+ (0.068)	0.282+ (0.072)
Gender (Male=1)		-0.125(0.425)	-0.065(0.682)
Training (Memory =1)		-0.799*** (0.000)	-0.762*** (0.000)
BRET		0.0008(0.870)	0.0012(0.806)
Economics			0.1133(0.485)
Age			-0.0596+ (0.096)
Number of experiments			-0.0069 (0.499)
Constant	-1.079 (0.143)	-2.295** (0.003)	-0.8463 (0.459)
Number of obs	2398	2358	2358
Number of groups	240	236	236
Log Likelihood	-1176.2113	-1140.2442	-1137.63

*** p≤0.001 ** p≤0.01 * p≤0.05 + p≤0.1 Table reports coefficients, p-values in parentheses.

Table 5b. Determinants of exploration (payoff)

Dependent variable: exploration	**Model 4**	**Model 5**	**Model 6**
Last payoff experienced	-0.002*** (0.000)	-0.0026*** (0.000)	-0.0026*** (0.000)
Aspiration-payoff gap (Negative gap=1)	1.099*** (0.000)	1.209*** (0.000)	1.207*** (0.000)
Payoff Slack	-0.0117*** (0.000)	-0.011*** (0.000)	-0.011*** (0.000)
Feedback (Yes=1)		0.366 (0.118)	0.328 (0.158)
Game (Numeric=1)		1.164*** (0.000)	1.136*** (0.000)
Gender (Male=1)		0.079(0.716)	0.1435(0.516)
Training (memory =1)		-0.536*(0.023)	-0.545+(0.054)
BRET		-0.003(0.659)	-0.003(0.647)
Economics			0.545* (0.016)
Age			-0.0872+ (0.078)
Number of experiments			-0.0077 (0.581)
Constant	3.618*** (0.000)	2.8718*** (0.000)	4.653*** (0.000)
Number of obs	2398	2358	2358
Number of groups	240	236	236
Log Likelihood	-1220.4995	-1178.9391	-1172.7362

*** p≤0.001 ** p≤0.01 * p≤0.05 + p≤0.1 Table reports coefficients, p-values in parentheses.

Table 6a. Determinants of exploitation of past routine (performance)

Dependent variable: exploitation of past routine	**Model 7**	**Model 8**	**Model 9**
Last performance experienced	-0.009+ (0.086)	-0.004 (0.443)	-0.0042 (0.418)
Aspiration-performance gap (Negative gap=1)	-3.957*** (0.000)	-3.899*** (0.000)	-3.895*** (0.000)
Performance Slack	0.0249(0.391)	-0.049+(0.053)	-0.0413+(0.099)
Feedback (Yes=1)		-0.612 (0.305)	-0.396 (0.510)
Game (Numeric=1)		-0.29(0.604)	-0.299(0.596)
Gender (Male=1)		-0.517(0.358)	-0.509(0.374)
Training (Memory =1)		6.538*** (0.000)	6.361*** (0.000)
BRET		0.003(0.874)	0.007(0.697)
Economics			-0.896(0.119)
Age			0.2278+(0.075)
Number of experiments			-0.047(0.208)
Constant	-3.046 (0.230)	1.129 (0.615)	-3.779 (0.309)
Number of obs	2398	2358	2358
Number of groups	240	236	236
Log Likelihood	-661.35367	-610.33674	-606.75329

*** p≤0.001 ** p≤0.01 * p≤0.05 + p≤0.1 Table reports coefficients, p-values in parentheses.

Table 6b. Determinants of exploitation of past routine (payoff)

Dependent variable: exploitation of past routine	**Model 10**	**Model 11**	**Model 12**
Last payoff experienced	-0.002*** (0.000)	-0.001* (0.02)	-0.0013* (0.017)
Aspiration-payoff gap (Negative gap=1)	-0.832*** (0.000)	-1.084*** (0.000)	-1.076*** (0.000)
Payoff Slack	-0. 003 (0.270)	-0. 01*** (0.000)	-0.009*** (0.001)
Feedback (Yes=1)		-0.312(0.522)	-0.2105(0.668)
Game (Numeric=1)		-0.007(0.987)	-0.0174(0.971)
Gender (Male=1)		-0.341(0.466)	-0.327(0.493)
Training (memory =1)		5.345*** (0.000)	5.239*** (0.000)
BRET		-0.004(0.809)	-0.001(0.946)
Economics			-0.779(0.103)
Age			0.113(0.267)
Number of experiments			-0.022(0.481)
Constant	-0.0259 (0.978)	0.309 (0.770)	-1.945 (0.439)
Number of obs	2398	2358	2358
Number of groups	240	236	236
Log Likelihood	-825.94047	-762.0587	-759.56457

*** p≤0.001 ** p≤0.01 * p≤0.05 + p≤0.1 Table reports coefficients, p-values in parentheses.

Table 7. Number of switches per clusters

Num. of switch. Clusters	0	1	2	3	4	5	6	7	8	9	1 0
1	27	35	36	23	1						
2			2	8	40	23	9				
3							10	11	10	4	1

Table 8. Composition of types

TYPE	Training (*Memory*)	Game (*Numeric*)	Gender (*Male*)	With feedback N=162	No feedback N=78
Others (N=66)	61 %	56 %	33 %	25 %	32 %
Consistent exploiters (N=29)	76 %	41 %	41 %	14 %	9 %
Unstable explorers (N=32)	38 %	75 %	44 %	12 %	17 %
Rational explorers (N=36)	50 %	53 %	64 %	14 %	17 %
Retreaters (N=23)	87 %	39 %	48 %	10 %	9 %
Partial exploiters (N=54)	15 %	35 %	48 %	25 %	17 %

Table 9. Descriptive statistics (only active phase)

	Type 0 Others (n = 659)	*Type 1 Consistent exploiters (n = 290)*	*Type 2 Unstable explorers (n = 320)*	*Type 3 Rational Explorers (n = 360)*	*Type 4 Retreaters (n = 230)*	*Type 5 Partial Exploiters (n = 539)*	*Post hoc[a]*
Performance*	**89.49**	95.43	94.44	90.25	94.39	<u>96.11</u>	0 < 1&2&4&5; 1 &2 > 3
Payoff*	403.19	374.65	484	461.36	**357.35**	<u>543.15</u>	0 &1&4 < 3 < 5 0 &1 &4 < 2 < 5 4<0
BRET*	31.57	<u>34.62</u>	31.47	32.03	**25.35**	28.98	4 < 0&1&2&3&5; 5<0&1&3 [b] 0<1 [b]

[*] p < 0.001 (Significance of mean difference; ANOVA)

[a] Post-hoc analysis based on pair-wise comparisons between types 0,1, 2,3, 4 and 5, (p <0.05)

[b] Post-hoc analysis based on pair-wise comparisons between types 0,1, 2,3, 4 and 5, (p <0.1)

Table 10. GLS Regression estimates

	Model 13 (Dep.variable: performance)	Model 14 (Dep.variable: performance)	Model 15 (Dep.variable: performance)	Model 16 (Dep.variable: payoff)	Model 17 (Dep.variable: payoff)	Model 18 (Dep.variable: payoff)
Consistent exploiters	5.927**	5.758**	5.9378**	-28.745	-12.061	-9.717
	(0.004)	(0.006)	(0.004)	(0.158)	(0.501)	(0.588)
Unstable explorers	4.941*	5.03*	4.725*	80.600***	55.164**	552.507**
	(0.013)	(0.013)	(0.021)	(0.000)	(0.002)	(0.003)
Rational explorers	0.747	0.844	0.7188	57.961***	48.893**	47.9216**
	(0.698)	(0.668)	(0.720)	(0.002)	(0.004)	(0.005)
Retreaters	4.888*	4.613*	4.523+	-46.05*	-22.468	-22.027
	(0.030)	(0.044)	(0.053)	(0.037)	(0.256)	(0.273)
Partial exploiters	6.601***	5.814**	5.527**	139.741***	111.659***	109.3722***
	(0.000)	(0.002)	(0.004)	(0.000)	(0.000)	(0.000)
Game(Numeric=1)		-0.258	-0.239		48.135***	47.667***
		(0.837)	(0.850)		(0.000)	(0.000)
Feedback (Yes=1)		4.13***	3.934**		35.768***	33.9206**
		(0.001)	(0.003)		(0.001)	(0.003)
Gender (Male=1)		-0.138	0.0637		1.991	3.539
		(0.911)	(0.960)		(0.853)	(0.746)
Training(Memory=1)		-0.915	-0.908		-78.422***	-77.9102***
		(0.502)	(0.507)		(0.000)	(0.000)
BRET		-0.0507	-0.0515		-0.539	-0.5439
		(0.200)	(0.198)		(0.114)	(0.115)
Economics			1.136			16.235
			(0.371)			(0.137)
Age			-0.3198			-2.7154
			(0.251)			(0.257)
Numbexperiments			0.0046			-0.0658
			(0.957)			(0.927)
Constant	89.503***	89.16***	95.639***	403.399***	417.023***	469.406***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of obs	2398	2358	2358	2398	2358	2358
Number of groups	240	236	236	240	236	236
Wald X^2	21.83**	36.13**	38.99	121.10**	246.71**	253.42
***R*^2**						
Within	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Between	0.0852	0.1382	0.1492	0.3409	0.5230	0.5330
Overall	0.0225	0.0371	0.0401	0.1309	0.2031	0.2070

*** p≤0.001 ** p≤0.01 * p≤0.05 + p≤0.1 Table reports coefficients, p-values in parentheses.

[1] Herbert Simon, in his seminal article (Simon 1955), suggested two alternative types of rationality: "global rationality" and "procedural rationality" (later known as bounded rationality). This distinction is elaborated on (Simon 1997, p. 17): "Global rationality, the rationality of neoclassical theory, assumes that the decision maker has a comprehensive, consistent utility function, knows all the alternatives that are available for choice, can compute the expected value of utility associated with each alternative, and chooses the alternative that maximizes expected

utility. Bounded rationality, a rationality that is consistent with our knowledge of actual human choice behavior, assumes that the decision maker must search for alternatives, has egregiously incomplete and inaccurate knowledge about the consequences of actions, and chooses actions that are expected to be satisfactory (attain targets while satisfying constraints)."

[2] These issues were already recognized by Simon (1955, p. 111, emphasis added): "The aspiration level, which defines a satisfactory alternative, may change from point to point in this sequence of trials. A vague principle would be that as the individual, in his **exploration** of alternatives, finds it easy to discover satisfactory alternatives, his aspiration level rises; as he finds it difficult to discover satisfactory alternatives, his aspiration level falls."

[3] We define an individual-level routine as a stabilized set of actions and choices that the decision maker can explain and teach to someone else (Laureiro-Martinez, 2014). An individual-level routine is thus a specific set of choices and actions by the individual with the aim to improve and stabilize task performance. In this sense our definition is compatible with the standard definition of organizational routines (Feldman and Pentland, 2003) in particular for what it regards the advantages – stabilizing and reducing costs – and the potential disadvantages – inducing rigidity and inertia triggered by the so called "competency trap" (March, 1991). Recent definitions of the concept of routine at the individual level, like the one suggested by Laureiro-Martinez (2014) who define the concept of routine within the frame of an individual tendency to routinized behaviors captures the essence of our definition. Quoting Laureiro-Martinez: "I define routinization propensity as «the individual tendency to develop and enact a behavioral repertoire that provides standard solutions (routines) for problems involving choice »" (p. 1112). Aligned with this view, the decision makers can decide to establish a repertoire of choices and consequent actions which increases their self-confidence in realizing an objective function (Giddens, 1991). It is worth noticing that this last motivation for the insurgence of a routine does not necessarily imply that the decision maker will develop an "optimal" routine. Nevertheless, we expect that the chosen behavioral repertoire would end with an improvement in performance, even with a satisficing behavioral assumption.

[4] The description of this process shown in Fig. 1 is a simplification of a real-life decision process as it implicitly assumes that one round of a new task is enough to learn the appropriate routine. We acknowledge this simplification for reasons of definitory parsimony and to avoid to complicate too much the interpretation of the experimental results. Furthermore, this definition includes the idea that from the first repetition onwards, a process of routinization takes place, even if "a routine" as such would not be completely established.

[5] In Table 1, we report the exact computer-screens visualized by participants during the experiment (note that they are in the original Italian language). Based on previous pilots, participants performed the tasks within a time limit calibrated to assure similarity of difficulty between the visual and numerical settings.