Modelling Bounded Rationality in Organizations: Progress and Prospects
Puranam, P; Stieglitz, N; OSMAN, M; Pillutla, M

© 2015 Taylor and Francis

For additional information about this publication click this link. http://qmro.qmul.ac.uk/xmlui/handle/123456789/10847

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact scholarlycommunications@qmul.ac.uk
Modelling Bounded Rationality in Organizations: Progress and Prospects

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Academy of Management Annals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>ANNALS-2013-0071.R4</td>
</tr>
<tr>
<td>Document Type:</td>
<td>Article</td>
</tr>
<tr>
<td>Keywords:</td>
<td>DECISION MAKING, LEARNING, RESEARCH METHODS, ORGANIZATION</td>
</tr>
</tbody>
</table>
Modelling bounded rationality in organizations: 

Progress and Prospects 

Phanish Puranam, Nils Stieglitz, Magda Osman, Madan M. Pillutla 

Abstract: 
Much of the formal modelling work in the organizational sciences relies on Herbert Simon's conception of bounded rationality, and it stakes a claim to drawing on behaviorally plausible assumptions about human behavior and action in organizations. The objectives of our review are three-fold. First, we summarize the formal literature by "model families" - classes of models sharing the same analytical structure- to highlight sharply the behavioral assumptions being made. Specifically, we discuss model families involving a) adaptation through search and learning by individual agents, b) mutual adjustment of interacting agents, and c) information aggregation in organizational decision-making. Second, we examine to what extent these models of bounded rationality in organizations are in fact consistent with the behavioral evidence in psychology and other related fields. Finally, we discuss opportunities for further research that strengthens the links between formal modelling in organizations research, and its behavioral foundations. In particular we highlight the promise of experimental methods that translate organizational models to multiple-subject experiments in the behavioral laboratory.

Note: 
An early draft of this article was prepared as a background note for a workshop on “The psychological foundations of organizational models” held at London Business School, May 11-12, 2012. We also thank Felipe Csaszar, Jerker Denrell, Dan Levinthal, and Thorbjørn Knudsen for helpful comments on earlier drafts of the paper. All remaining errors are ours.
1. Introduction

The work of Herbert Simon (1916-2001), his colleagues and others inspired by his vision of constructing a science of organization has always had a strong component of formal (mathematical and computational) modelling (e.g. Simon, 1955; Cyert and March, 1963).

Formal models allow for a precise statement of assumptions in a manner that makes them transparent, and allows for an examination of the possibly surprising ways in which a system characterized by these assumptions behaves over time. Verbal theorizing does not offer either this precision or insight into dynamics, unless practiced by an extraordinarily gifted theorist (Lave and March, 1975; Davis et al., 2007; Harrison et al. 2007; Adner et al., 2009).

The tradition of modelling behavior in and of organizations inspired by Simon and his colleagues stands in contrast to other forms of modelling in the social sciences primarily because of the centrality of the notions of bounded rationality.

Simon’s (1955) notion of bounded rationality assumed that 1) the decision making process begins with the search for alternatives 2) the decision maker has egregiously incomplete and inaccurate knowledge about the consequences of actions, and 3) she chooses actions that are expected to be satisficing (attain targets while satisfying constraints). Later work by Cyert and March (1963) added an important elaboration - problemistic search- which indicates that the search process described above is itself triggered by a failure to meet aspiration levels (see also recent overviews by Gavetti et al (2007; 2012) of the tradition of research inspired by Simon). We will refer to this form of bounded rationality, to paraphrase Simon, as a form of “adaptive rationality” as it emphasizes the process of searching for better alternatives by reacting to feedback (e.g. Simon, 1996).  

---

1 Simon’s conception of global rationality was as follows (Simon, 1997:25): “This is what I call full or global rationality: people are making their decisions to maximize utility in a world which they either understand exactly or in terms of a known probability distribution (i.e. they are maximizing subjective expected utility).” Since bounded rationality is defined by its departure from global rationality, it follows that there are many possible ways in which rationality can be bounded. There is, of course, a significant body of modelling work with some form of bounded rationality in economics; see for instance Rubinstein (1998) or Spiegler (2011), as well as an extensive empirical tradition (e.g.
Simon’s intention was to theorize about and to model organizations in a manner that took account of the nature of the human agents who constituted them. Specifically, if human agents are best characterized as being bounded, rather than globally rational when confronted with typical problems in organizational settings, then we should model organizations with these ‘correct’ behavioral assumptions since organizations are aggregations of these human agents. It is not the purpose of this essay to review the evidence for the proposition that individuals are boundedly rather than unlimitedly rational, which we take as granted. Instead, our focus is on investigating whether models of organizations that assume boundedly rational agents make psychologically ‘correct’ assumptions about humans—in other words to assess their behavioral plausibility.

Since Cyert and March (1963), about a 100 papers invoking models of bounded, adaptively rational behavior in and by organizations have been published in the leading outlets for research about organization and management theory. A common feature of these papers is their explicit declaration of building on behaviorally plausible models of human action—i.e. on realistic assumptions about human behavior in organizations. A few have achieved canonical status such as the “garbage can” model of Cohen, March and Olsen (1972) and March’s model of exploration and exploitation in organizational learning (1991). These models and their insights have percolated widely beyond the community of modellers in the organization sciences to an extent where it is no longer necessary to make them the focus of review articles.

---

2 The statement by Schoemaker (1982) in his review of the evidence on expected utility (EU) maximization remains a good summary: “For well-structured repetitive tasks, with important stakes, and well-trained decision makers, EU maximization may well describe the actual decision process.” In other words unbounded rational choice may be a useful theory when there is a small or well-known space of choices, (expected) utilities associated with the choices have been discovered and are now...
such as this one (there is an excellent one by Gavetti, Greve, Levinthal, & Ocasio (2012) in this very journal for those interested in the research inspired by these classic models from the Carnegie tradition). Rather, our objective in this essay is to do three things not undertaken so far:

**First**, as noted, we wish to examine the growing body of modelling work based on adaptive rationality in order to examine to what extent these models of organizations are in fact consistent in their assumptions about behavior with the available scientific evidence. If this consistency is a central claim for the distinctiveness of modelling by organization and management theorists, then it would be good for the field to explicitly evaluate the strength of these claims. Drawing on assumptions about human behavior that are better grounded in psychology than traditional rational choice models is an appealing feature of these models for management and organizations researchers that sometimes remains obscured behind the technical details of the models. By exposing this, we aim to increase the appreciation for formal models among organizations researchers, which as Adner et al. (2009) point out is still fairly limited.

**Second**, we will systematically summarize this body of work by “model families” - classes of models that share an underlying analytical structure, with a distinctive set of assumptions and mechanisms, and resulting insights. Specifically, we discuss model families which examine a) individual agent adaptation through search and learning, b) mutual adjustment of an organization’s constituent agents and c) information aggregation. Such an exercise we believe has enormous instructional value for modelers as well as organizations researchers in general, by making transparent underlying assumptions, as well as the kinds of problems to which the models may bring insight, and finally by suggesting avenues for empirical research.

well known, and there is sufficient difference between the outcomes of choices to make maximization easy. Such situations, Simon argued, are rare within organizations.
Third, we discuss opportunities for further research that strengthen the links between formal modelling in organizations research, and its behavioral foundations. Surprisingly, direct tests of the empirical predictions of organizational models of adaptive rationality have been rare, with these models more often being treated as metaphors that shed light on how previously known mechanisms might interact in surprising ways. While this is valuable, what makes models most useful is their ability to predict. Beyond the usual empirical approaches management researchers favour (e.g. large sample archival data analysis, surveys, case studies), in our view, there is also an opportunity to use experimental methods to test model predictions. These experiments could translate agent-based models to multiple-subject experiments in the behavioral lab as a methodology to test the predictions. Pragmatically, since both modeling and experimental work deal with interaction structures among a small number of agents, the possibilities for theory building, calibration and theory testing are numerous at their interface. Empiricists may benefit from sharper statements of hypotheses that model based predictions can generate, while modelers may benefit from the data that empiricists gather that can help falsify current models and force them to improve in their veridicality and usefulness.

Thus, at a broader level, our goal in this article is to bridge the chasm between empirical researchers (including experimentalists) and theoretical researchers in order to help stimulate the production of scientific (i.e. behaviorally grounded) models in organization and management theory. Our efforts are guided by the words of the Nobel Laureate Vernon Smith (1982: 924) who argued that ‘theory should be ever more demanding of our empirical resources. Simultaneously, data should be ever more demanding of the empirical relevance of theory and of the theorist's expertise in working imaginatively on problems of the world, rather than on stylized problems of the imagination’. The modeling work on adaptive rationality fulfills Vernon Smith’s criterion of theory that captures problems of the world as
its avowed purpose is to model the important problem of ‘organizing.’ It is therefore
deserving of empirical testing and we hope that our paper stimulates that.

The rest of this article is organized as follows: in section 2, we describe the common
building blocks that are present in all models of bounded adaptive rationality. Section 3 is the
heart of the paper, and examines each of three important model families that appear often in
organizations research- models of individual adaptation, mutual adjustment, and information
aggregation. For each model family, we describe the model’s structure, discuss the extent to
which the assumptions of the model have sound foundations in the evidence about how
individuals and groups behave, and then state the key insights emerging from the model. We
highlight the core results that emerge for each model family rather than review the specific
results that are derived in each paper within this literature; this is because the results in
particular papers point to “moderators” that weaken or strengthen the effects underlying core
results. This aggregation into model families also implies that we are unable to do justice to
every influential model (e.g. Carley, 1992; Prietula, Carley, Glasser, 1998; Davis, Eisenhardt,
Bingham, 2009), as we pick those which have underlying platforms that are shared across
multiple papers and authors. Finally, section 4 concludes with a discussion of future research
opportunities.

To preview the fruits of this review, (which are discussed more fully in the last section), we a)
identify the four key behavioral assumptions that underlie the models we have reviewed,
namely exploratory choice, learning through reinforcement, search through local hill
climbing, and problematic search, and verify whether these assumptions have an empirical
basis b) discuss testable implications from these models that have not so far been examined
by empirical researchers, and provide a useful mapping between model parameters and
empirically observable data in Table 1, and c) discuss the possibilities of using behavioral
laboratory experiments to test models of adaptive rationality.
2. Basic elements of a model of bounded adaptive rationality

Before moving to a discussion of different model families, we briefly describe the foundational elements of models of boundedly adaptive rational behavior.

1. **Task Environment**

This describes an objective reality which can be characterized with respect to an agent in terms of a set of possible *Actions* \( \{A_1, A_2, \ldots, A_m\} \) which map into *Performance outcomes* \( \{\Pi_1, \Pi_2, \ldots, \Pi_m\} \) for the agent. The task environment need not be stable; it could contain other agents, whose behavior may itself contribute to the instability of the task environment. Implicitly, given a task environment, the objective of the agent in an adaptive rationality model is to take actions that yield a level of performance deemed acceptable (often termed the aspiration level). Task environments thus capture a mapping between the possible courses of action available to an agent, and the likelihood of attaining the desired goal of the agent for each course of action. Task environments are tied to particular goals; a change in goals will imply a change in task environments.

2. **A Representation** of the Task Environment

An agent at any point in time may have beliefs (not necessarily knowledge) of the *Task Environment*, in terms of two things:

i. a set of feasible actions \( \{a_{1,t}, a_{2,t}, \ldots, a_{m,t}\} \),

ii. and the associated set of performance outcomes \( \{\pi_{1,t}, \pi_{2,t}, \ldots, \pi_{m,t}\} \).

3. **A Choice process** *given* the Representation

This is a procedure by which the agent chooses an action within his *Representation.*
These three elements of a model of boundedly rational behavior correspond to the three hierarchically linked components that have been used to characterize behavior in classical organization theory (e.g. Simon, 1946; Cyert and March, 1963): Goals (operationalized through the task environment), Representations and Choice processes. Interestingly, these elements are present in rational choice models as well. The difference is that the task environment in a rational choice model is typically assumed to be known perfectly to the agent so that representation and task environment are identical (or at least the former is a known and unbiased probability distribution over the latter), and the choice process is the maximization of subjective expected utility. To make the modelling of the choice process mathematically tractable, it is also common in many such models to assume certain mathematical properties of the task environment that make it easy to apply standard techniques to find the best action.

Bounded adaptive rationality models are different. The first (and in our view both necessary and sufficient) point of distinctiveness for such models is that the representation is incomplete and/or incorrect in terms of the set of possible actions as well as their associated outcomes. The agent ‘has egregiously incomplete and inaccurate knowledge about the consequences of actions’, to paraphrase Simon. Thus at any point in time, the agents Representation of actions \( \{a_{1,t}, a_{2,t}, \ldots, a_{k,t}\} \) may contain only a small subset of actions from \( \{A_1, A_2, \ldots, A_m\} \) (i.e they have a “coarse” representation), and the beliefs about associated outcomes \( \{\pi_{1,t}, \pi_{2,t}, \ldots, \pi_{k,t}\} \) may be incomplete, inaccurate or both with respect to the actual values of outcomes in the Task Environment \( \{\Pi_1, \Pi_2, \ldots, \Pi_m\} \).

A second point of departure (which, in our view is sufficient but not necessary to make it a model of bounded rationality) is that the Choice process may not involve maximization even \textit{within} the representation. More specifically, alternatives to maximization include the notion
of “non-greedy” or exploratory choice (Sutton and Barto, 1998), whereby an agent may with some probability select an action other than the one that is associated in the Representation with the best outcome. A related concept is “satisficing” which involves making a choice relative to an aspirational level (Cyert and March, 1963; Greve, 2003), rather than pursuing a course of action aimed at the best possible outcome.

4. Critical to a model of bounded but adaptive rationality is a Transformation procedure that modifies Representations and possibly Choice processes over time. The transformation process may be a consequence of communication with other agents, a process of learning- a change in beliefs as a consequence of experience- or indeed random fluctuations in beliefs. The process of Transformation may itself be triggered by the consequences of the Choice process, in the form of feedback from the Task Environment- the realized outcome. It may affect not only Representations but also Choice processes by changing aspiration levels. There is usually at least a tendency for the Transformation process in models of adaptive rationality to move Representations towards alignment with Task Environments, but this can take time, and is seldom either complete or comprehensive.³

Thus the typical structure of an adaptive rationality model involves some or all of the following (possibly) recursive sequence:

1. Agent has (initial) Representation of the Task Environment.
2. Agent selects an action through a Choice process.
3. Feedback is generated from the consequences of the agents action in the Task Environment.

³ A process of Selection (through competition) sometimes accompanies organizational models of adaptive rationality (e.g. Levinthal, 1997; Levinthal and Posen, 2007), and may be an alternative (or addition) to the Transformation process in terms of generating change over time. Put simply, the individual agents in the model may change their representations through Transformation processes, or the distribution of representations in the population may change through Selecting out those agents who have poor representations, or both. Since selection through competition does not feature in all models of adaptive rationality, we only discuss it when relevant.
4. A *Transformation* process of the *Representation* and possibly *Choice* process is triggered by the feedback.

5. (Step 1 repeats).

As should be clear, our focus is on models that not only feature bounded rationality (i.e., in which agent’s representations are imperfect) but are also adaptive (i.e. there is a transformation process that modified representations over time).\(^4\)

Adaptively rational models of organizations vary in whether they model organizations as ‘unitary’ actors\(^5\) or as interacting agents whose sometimes conflicting preferences, beliefs and choice procedures cannot be captured conveniently in aggregate form as if they relate to a single agent. This suggests that we must consider two approaches to assessing behavioral plausibility. First, to what extent do unitary actor models of organizations embody assumptions about their adaptive rationality that are consistent with how groups and organizations as aggregations act? Specifically, there is much research in social psychology and organizational behavior about how and when groups and other collectives act (as an aggregate), and indeed act differently compared to individual actors in similar situations. We use that research as the basis to verify the validity of the behavioral assumptions of models of adaptive rationality that employ unitary actor assumptions. The second approach to assessing behavioral plausibility is to see how multi-agent models of organizations embody assumptions about the agents’ adaptive rationality that are consistent with how individuals act.

\(^4\) Models with the first feature but not the second also exist in the organizational literature, and often invoke maximization within an imperfect representation (e.g. Siggelkow, 2002; Kretschmer and Puranam, 2008; Puranam and Vanneste, 2009; Gulati and Puranam, 2009). Since there is considerable variation in the nature of the Task Environments and Representations in these papers, we do not attempt to summarize these here.

\(^5\) The unitary actor assumption can be justified if there are rules of preference/belief aggregation, and selection into or out of an organization does not destabilize such processes; if an organization is effectively controlled by a powerful individual, or a governing coalition which has stable preferences, beliefs and choice procedures (see for instance Hug (1999) for a formal analysis of the conditions under which unitary actor assumptions may be more or less plausible).
3. Families of organizational models of adaptive rationality

We summarize the modelling literature on adaptive bounded rationality in organizations in terms of three main model families. The first, featuring models of individual agent adaptation, provides some or all of the building blocks for the second (mutual adaptation) and third (information aggregation). The literature is also most extensively developed around the first. Accordingly, the coverage of models of individual agent adaptation – particularly in terms of a discussion of their behavioral plausibility - is more extensive in this review than of the other two model families. It is our hope that the field will develop in a way that allows future reviews, such as this one, to give more coverage to the second and third families of models, which are very relevant to issues of aggregation in organizations.

For each of the models that we present, we begin with some information about the conceptual background of the model and, present its core elements (task environment, representation, choice, transformation of representation) to identify the behavioral assumptions made in the model. We then examine the behavioral plausibility of core assumptions and move on to discussing the key insights from the model.

A. Models of Individual Agent Adaptation

Since the discovery of superior (to status quo) alternatives through an adaptation process is the essence of adaptive rationality, we begin with a look at models of individual agent adaptation. We focus on models in which decision-making by an agent plays an explicit and central role in the model. Two important sub-classes of organizational models are considered within this family. The multi-armed bandit depicts learning and decision-making under uncertainty (and models organizational problems such as resource allocation, including but not restricted to making decisions about different investment opportunities, allocation of resources to different divisions of a company, etc.), while the NK model examines adaptation
in complex task environments that feature interdependencies between choices (capturing organizational problems such as the formulation of new business models, adoption of manufacturing or human resource policies with unknown interactions between them).

**a. Reinforcement learning in multi-armed bandit models**

This model takes its name from the slot machines in gambling establishments (Brand, Wood, & Sakoda, 1956; Brand, Sakoda, & Woods, 1957). Metaphorically, the model describes a gambler who faces a slot machine with multiple alternative arms, each with an unknown payoff distribution. The gambler has a fixed budget and intends to make as much money as possible over multiple trials. She learns about the payoff distributions associated with an arm by selecting it and observing the realized payoff. Obviously, the gambler always wants to select the alternative with the highest expected payoff, but the problem is that she does not initially know the expected payoffs for each arm, and can only learn them by actually trying each of the multiple arms. Thus the gambler’s behavior is a process of repeated trials, followed by learning from the outcome of each trial.

Lave and March (1975) and March (1996) introduced the bandit formulation to organizations researchers; subsequent formal papers include Denrell and March (2001), Posen and Levinthål (2012), Stiglitz, Knudsen, and Becker (2014), Puranam and Swamy (Forthcoming) and Lee and Puranam (Forthcoming).

Modellers in the adaptive rationality tradition have made use of the multi-arm bandit’s strong structural similarity to resource allocation processes in organizations: the model describes with reasonable fidelity a decision maker who aims to maximize returns from her investments, and is confronted with multiple alternatives with unknown performance consequences. The relative attractiveness of these alternatives can only be gauged by actually trying them (i.e., making some investments) and observing and interpreting the resulting feedback. The goal is
to discover the alternative with the best returns as quickly as possible. In an organizational context, the uncertain investment alternatives could, for example, represent different entrepreneurial opportunities (Minniti and Bygrave, 2001; Gruber et al., 2008), different divisions in a multi-product firm (Noda and Bower, 1996), or different research projects (Loch and Kavadias, 2002).

We describe in more detail below the four key elements of the multi-armed bandit model before examining the behavioral plausibility of core assumptions and the key insights from the model.

1. Task Environment

The task environment in a bandit model consists of a set of ‘m’ discrete alternatives, each representing a possible course of action with distinct payoff consequences. Each alternative is associated with a unique payoff distribution (whose properties are unknown to the agent). This a) creates performance differences among alternatives and b) makes performance feedback noisy and potentially misleading. False negatives may arise – a good alternative looks bad because of an atypical unlucky draw from its payoff distribution, as can false positives – a bad alternative may look good because of an atypical lucky draw from its payoff distribution.  

To illustrate these ideas, consider a bandit with m=2, with alternatives A and B. A simple everyday example is finding a good place to eat, with A and B representing two restaurants. One alternative is in an objective sense superior for the decision-maker: alternative A

---

6 The basic bandit setup usually assumes a stable task environment, i.e., the probability distribution of payoffs for each arm is fixed. However, the model can be easily amended to also consider dynamic task environments where the probability distributions fluctuate over time to depict changing markets or technological conditions for the opportunities that each alternative represents (Posen and Levinthal, 2012, Stieglitz et al., 2014). Similarly, competence development and lock-in to particular alternatives may be represented by making the probability distribution conditional on the number of times an alternative has been chosen (Denrell and March, 2001).
dispenses a reward (of 1) with a 0.75 probability, while alternative B only has a success rate of 0.5. In the restaurant example, a payoff of 1 would represent good food and a payoff of 0 - bad food. A fully informed decision-maker would therefore always pick alternative A. The bandit task is trivial as long as agents have full information about the alternatives. However the purpose of the model is to describe the process that ensues when the knowledge about alternatives – the agent’s representation (i.e., in the above example, they do not know the probability of a reward associated with each alternative) – is far from complete initially.

2. Representation:

Commonly, in bandit models, the set of alternative actions in the agent’s representation is complete and correct with respect to the Task Environment, but the associated performance outcomes are not. This corresponds to situations where the space of alternatives is known, but not their value. Thus the CEO of a multi-business firm may know the number of divisions in the firm that need capital in the budgeting process, but not the true expected rates of return for investing in them. Similarly, the number and location of restaurants may be known, but not their quality. Instead, for each action in the agent’s representation, the agent may have at any point in time (possibly egregiously inaccurate) beliefs (Yechiam and Busemeyer, 2005; Posen and Levinthal, 2012). In the technical literature these are referred to as “action value estimates” (Sutton and Barto, 1998). For example, the action value estimates for alternatives A and B could be 0.5 and 0.66 respectively in the example above. Assumptions about these prior beliefs represent the current state of knowledge of the agent, e.g., she could have heard from friends that restaurant B is much better than A.

3. Choice process:

Given the agent’s representation, there are broadly two kinds of choice rules: maximizing and probabilistic choice rules (Yechiam and Busemeyer, 2005). Maximizing choice rules assume that agent always picks the alternative currently associated with the highest expected payoff in the agent’s representation. This is also sometimes called greedy action selection (Sutton
and Barto, 1998). For example, an agent following a maximizing choice rule would select alternative B based on her priors in our example, which is the incorrect choice. In the literature, such a choice would be termed an exploitative one, as it picks the alternative considered best within the current representation.

Probabilistic choice rules, in contrast, assume that choices are probabilistically determined in proportion to the current relative strengths of the representations. A common probabilistic choice rule in organizational models is the so called “softmax” action selection rule (Luce, 1959; Sutton and Barto, 1998; Posen and Levinthal, 2012). In this rule the probability of a choice k at time t depends on its relative action value estimate V in the agent’s representation and an exploration parameter $\tau$.

$$p(k)_t = \frac{e^{V_{a_t}} / \tau}{\sum_{b=1}^{m} e^{V_{a_b}} / \tau}$$

The parameter $\tau$ tunes the degree of exploration – the probability of choosing an alternative currently seen as inferior within the representation – by discounting the current representations. A higher $\tau$ discounts current representations more strongly and thereby increases the probability of selecting a (currently believed to be) inferior action. For example, given the (wrong) representations for A (0.5) and B (0.66), the probability of choosing B with a very low degree of exploration $\tau = 0.1$ is 0.75. In contrast, with $\tau = 1$, the probability of choosing B is just 0.52. Note that there is a clear benefit to exploration here, because the agent currently holds a flawed representation, believing wrongly that alternative B is superior.

For a given task environment, Gittins and Jones (1974) formally proved the existence of an optimal level of exploration in the bandit problem. However, as is generally acknowledged, the information and computational power required to implement this optimal exploration level makes this intractable except under very particular assumptions (Gittins, 1979; Arthur, 1991; Posen and Levinthal, 2012). Organizations research inspired by these models has tended to
use either constant levels of exploration, comparing across levels (in order to study the effects of different levels of exploration on system performance), or simple behaviorally plausible rules for deciding the level of exploration dynamically (in order to focus on other parameters of interest in the model). For instance, empirical work in the behavioral theory of the firm tradition argues that organizations increase exploration as a response to underperformance relative to an aspiration level (e.g. Greve, 2003; Hu et al., 2011). Applied to the softmax, this would mean that parameter $\tau$ is contingent on performance relative to aspiration, with a higher $\tau$ if the organization underperforms relative to its aspiration.

4. Transformation process:

Since most bandit models assume that the agent’s representation is complete in terms of the action space, but is incorrect ex ante in terms of the associated payoffs, the transformation processes of interest in such models involve how the representation of the task environment changes over time. Agents typically engage in reinforcement learning—whose variants are known under the labels of trial-and-error learning, experiential learning, operant or instrumental conditioning, “win-stay-lose-shift” rules in the relevant literatures in psychology, computer science, organization theory and evolutionary biology (Thorndike, 1911; March, 1991; Domjan, 2010; Nowak and Sigmund, 1993; Sutton and Barto, 1998). All instantiate Thorndike’s Law of Effect: “Responses that produce a satisfying effect in a particular situation become more likely to occur again in that situation and responses that produce a discomforting effect become less likely to occur again in that situation.” Thorndike’s law of effect implies that favourable feedback tend to positively reinforce the belief about an alternative and thereby make its choice more likely.

The most basic way to model the reinforcement learning process is a simple averaging rule, where the representation simply reflects the average of all payoff signals received in the past. For example, assume that an agent visited restaurant A four times and received the following
payoffs [1,0,1,0]. Taking the average leads to a belief of 0.5. Restaurant B was selected six times, with payoffs [1,1,1,0,0,1], so the belief in the agent’s representation is 0.66.

A more general approach to agent’s reinforcement learning is based on statistical learning models (Bush and Mosteller, 1955; Denrell and March, 2001) that explicitly consider a learning rate. For example, the exponential recency weighted average rule discounts rewards that occurred long ago:

\[ V_{m,t} = V_{m,t-1} + \phi [r_{m,t} - V_{m,t-1}] \]

Where \( \phi \in [0,1] \) is called the step size parameter (Sutton and Barto, 1998) and can be interpreted as a learning rate. The higher the parameter \( \phi \), the more recent rewards matter in the representation, and the more rapid is the updating process. For instance, assume that the current representation of B is 0.66 and the agent selects B and receives no payoff. With \( \phi = 0.9 \), the new representation is 0.066, a massive re-evaluation of the alternative. In contrast, with \( \phi = 0.1 \) the new representation is 0.594, a much more modest change in representation.

The bandit model can also be extended by considering competition based selection in a population of agents, in which the relative cumulative payoffs influence survival probability (Denrell and March, 2001; Stieglitz et al., 2014). Some recent organizational models have begun to look into a decision-making context in which actions are not followed by immediate feedback (Denrell, Fang & Levinthal, 2004; Rahmandad, 2008; Fang and Levinthal, 2009). Rather, feedback may only be available after a sequence of actions has been performed. The delayed feedback conditions relate to the notion of credit assignment, i.e., the problem of assigning payoffs to alternative actions (Fang, 2012). The modelling of learning in this context is also based on reinforcement learning principles in which action value estimates are assigned to intermediate actions.

**Behavioral plausibility of bandit models**
There are two key behavioral assumptions about the agent in a bandit model: **reinforcement learning** and **exploration in choice**.

There is an extensive body of experimental evidence documenting that reinforcement learning is a good model for the behavior of individual adaptation in situations where the initial understanding of a complex task environment is limited, (e.g. Erev and Roth, 1998; Camerer, 2003; Domjan, 2010). The fundamental validity of the Law of Effect in such situations is beyond dispute (Thorndike, 1911; Domjan, 2010; Nowak and Sigmund, 1993). Indeed, there is evidence suggesting that as the computational load of a task that an agent is performing increases, they are more likely to rely on reinforcement learning rather than alternative heuristic type strategies in choice tasks (Otto et al, 2010).

Experimental evidence also suggests that the softmax choice rule – which implies explorative choice- is a good approximation to the neurological processes that underlie how humans select actions in trial and error learning situations, in that exploration occurs, and does so through inhibiting tendencies towards greedy (exploitative) action selection (Camerer and Ho, 1999; Daw et al, 2006; Laureiro-Martinez, Brusoni, Canessa, & Zollo, 2014). However, the conditions under which people switch between exploration and exploitation are not yet fully understood. Some recent studies have shown that people appear to enter uncertain choice situations by exploring (Lea et al, 2011; Steyvers et al, 2009), and the rate of switching back and forth between exploration and exploitation decreases as a function of experience.

Notwithstanding these broad consistencies, the parameters of the learning process, such as exploration and learning rates, differ significantly between people. Recent research has focused attention on the sources of these differences, using psychometric measures of cognitive ability (Steyvers, et al, 2009), and neurological differences (Badre, Doll, Long, & Frank, 2012), as well as differences based on dispositions to seek or avoid novelty (Payzan-LeNestour, & Bossaerts, 2012). These findings provide empirical opportunities as the
individual differences can be measured and used as proxies (for parameters of the learning
process) to investigate whether predictions about how variations in the parameters of the
learning process produce expected results when coupled with particular properties of the task
environment.

In many organizational applications, it is the organizational unit as a whole that is assumed to
undertake reinforcement learning, with exploration. One way to justify this is to appeal to the
notion of a powerful key actor as representing the organization. In this case, the results at the
individual level can be directly used to justify the modelling assumptions.

The other is to impute learning behavior to the organizational level on the basis of
hypothesized internal dynamics such as the strengthening of coalitions, or the creation of
routines as Cyert and March (1963), and later Nelson and Winter (1982) did. There is a
significant body of work that indirectly documents organizational level reinforcement
learning through the phenomenon of organizational experience curves (e.g. Argote, 2012;
Zollo and Winter, 2002). These depict an improvement in performance with experience, but
with the rate of increase slowing down with experience. Since this pattern of performance
improvement with experience is a hallmark of reinforcement learning processes, the former
provides indirect evidence for the latter.

How organizations balance exploration and exploitation has also been an active area of
research since March (1991) first introduced the trade-off to the organizational literature.
Under the term “ambidexterity” (O’Reilly and Tushman, 2013) a significant empirical
literature now exists that includes cases as well as some large sample evidence on how
organizations attempt to balance exploration and exploitation- leaving little doubt that the
attempt is frequently and deliberately made, even though is not always successful (Brown and
Eisenhardt, 1997; Raisch et al., (2009), and also Gavetti et al., (2012) for a recent review of
the literature inspired by the behavioral theory of the firm).
In sum, it would appear the behavioral principles that underlie multi-armed bandit models - such as reinforcement learning and explorative choices - rest on reasonable empirical foundations, whether used to model individual or organizational behavior in contexts where experience is necessary to generate information about the value of alternatives. As the research in psychology and organizations converges towards a better understanding of when exploration is triggered, it may be possible to incorporate these refinements into the next generation of models.

**Key trade-offs and insights**

Having examined the behavioral plausibility of its assumptions, we now turn to the key underlying trade-offs and insights generated by this family of models. These insights are worth investigating empirically. There are two key trade-offs that arise in the bandit model. First, there is the canonical trade-off between exploitation of current knowledge and exploration of new knowledge (Holland, 1975): The agent may rely on current knowledge by selecting the arm known to produce a good outcome (exploitation) or choose an apparently inferior arm in the hope of attaining a superior outcome and thereby improving future payoffs (exploration). Because initial representations of agents usually do not correspond to the task environment there is a positive value of exploration because it prevents agents from settling on an inferior alternative. However, agents may also overinvest in exploration – that is, they direct far too much effort on choosing inferior alternatives, without experiencing substantial benefits that will improve their future (Fernie, & Tunney, 2006; Fridberg et al. 2010; Kjome et al., 2010; Premkumar et al., 2008; Steingroever et al., 2012; Wood et al., 2005).

Figure 1 shows in a stylized manner the standard result on the value of exploration in a multi-armed bandit model with a stable task environment. On the vertical axis, the performance of agents is reported. Note that the performance measure used here is the cumulative payoffs over all trials. This represents settings in which agents face persistent selection pressures.
(Denrell and March, 2001; Levinthal and Posen, 2007). One could also simply report performance in the n-th trial, or the fraction of cases in which the best alternative has been found by trial n. The horizontal axis plots different values of the exploration parameter in the softmax choice rule, with low values corresponding to low levels of exploration. Figure 1 shows that agents underinvesting in exploration suffer performance shortfalls. They do not explore enough and thereby settle on an inferior alternative. To pick up again on the restaurant example, an agent settling for the first restaurant with decent food as her favourite place misses out on many superior establishments. In contrast, agents with a very high exploration level spend too much on exploring inferior alternatives. That is, she visits and spends money on many bad restaurants again and again in the hope that things improve. Agents that strike exactly the right balance between exploration and exploitation show superior performance because they identify the best alternative (by providing for sufficient exploration) and firmly hold on to it (by providing for enough exploitation).

![Figure 1: The optimal balance between exploitation and exploration](image)

All models using the multi-armed bandit implicitly or explicitly feature the exploration-exploitation trade-off, and some may explicitly examine moderators that strengthen or weaken the forces generating this trade-off. For instance, an important paper that has analysed the location of the optimal level of exploration as a function of turbulence in the task environment is Posen and Levinthal (2012). They show that, optimally exploration should
decline as the possibility of changes to the payoff distribution for the arms in the bandit
increases to high levels.

The second trade-off has to do with the potential pitfalls of learning too rapidly in
environments with noisy feedback (Denrell, 2007). Because feedback is noisy – payoffs are
drawn from a probability distribution – the agent is confronted by “false-positives” and
“false-negatives”, i.e., when the value of an action is believed to be higher than it really is and
vice versa. Clearly some learning is better than no learning at all, but as several organizational
models point out, rapid learning is more susceptible to superstitious learning and mistakes, in
which the agent may, select the wrong alternative by learning too rapidly from false-positives
and false-negatives. The agent can superstitiously associate purely chance outcomes with her
own chosen actions (Blanco, Matute, and Vadillo, 2011; Lave and March, 1975). (Consider
the proverbial “lucky socks” which get associated with positive outcomes by an individual
who happened to wear them to many good outcome situations).

Papers that have examined the consequence of this learning problem show that agents can
learn to become risk averse over time (March 1996; Denrell and March, 2001) and thereby
may systematically underinvest into higher-variance, more innovative outcomes. The related
“hot stove” effect (Denrell and March, 2001) identifies an important asymmetry between false
positive and false negative learning: agents will stop selecting alternatives associated with
false-negative outcomes and thereby may miss out on superior alternatives, whereas false
positives will likely be corrected for as the falsely attractive alternative is tried again. Thus a
bad restaurant that generates a lucky good meal (false positive) will be found out on being
visited again, but a good restaurant that produces an unlucky bad meal (false negative), may
never be visited again.

Figure 2 shows this second trade-off in a stylized manner. The vertical axis gives the
probability of having found the best alternative after a number of trials (obtained by
computing the fraction of agents in a simulation who have chosen the correct alternative), and the horizontal axis shows increasing rates of learning. As the figure shows, in noisy environments, rapid learning can do worse than slower learning.

![Figure 2: Learning rates and performance in different task environments](image)

The insights from the bandit model for the designers of learning systems, including organizations, can be inferred from these trade-offs. The parameters of the learning process crucially determine success when entering poorly understood and noisy task environments. In particular, avoiding the extremes of very low or very high exploration, as well as very low or very high learning rates, can help promote valid learning and successful choice of alternatives.

The trade-offs and insights generated by the multi-arm bandit problem illustrate the point we made at the beginning that formal models have the ability to generate systematic insight in ways that verbal theorizing may not be able to. As we will note in the final section, the

---

7 A moderate level of exploration is assumed. At very high levels of exploration, the effects of learning are suppressed, since choices do not discriminate among alternatives, regardless of what has been learned about them.
precision of these predictions suggests the need and appropriateness of laboratory methods for
testing them.

b. Models of search on rugged landscapes

In addition to the multi-arm bandit problem, modellers examine individual learning and
adaptation using the so-called ‘NK’ models. These models examine organizational decision
making problems that are less structured than the ones that multi-arm bandit problems
examine. Decisions about new products or processes, new business strategies, or new
organizational forms and designs that are often attributed to a recombination of existing
resources, technologies, or design elements (Fleming, 2001; Denrell et al., 2003; Lippman
and Rumelt, 2003) are examples of such problems. NK models differ from the bandit
problems in three important ways. First, unlike bandit problems, the NK models do not
assume that the set of possible actions is given and known by the agent. That is, the
representation of the task environment does not include all possible actions. Second, actions
that differ in payoffs are not equally spaced from one another. Third, in choice situations with
multiple pay-off relevant elements, the best option on one element often depends on choices
made about other elements. In other words, there is interdependence between different choice
elements.

The NK model was originally developed in evolutionary biology (Wright, 1932; Kauffman
and Weinberger, 1989; Kauffman, 1993). Levinthal (1997) introduced the model to the
organizations and management literature to study the interrelationship between adaptation and
selection of organizational forms. The model takes its name from the two key parameters of
the model that establish the task environment: The N parameter of the model represents the
performance-relevant choice elements, while K is a parameter influencing the
interdependencies among these elements.
An everyday example helps to provide the key intuition behind most NK models. Suppose the
design of a car involves choices about four key design elements (the N in the model): engine
size, overall shape, placement of trunk and colour. If there are three choices for each design
element then the space of actions contains 3 X 3 X 3 X 3 = 81 points, each of which has an
associated performance level (this could be willingness to pay for an average car buyer). If
the best choice on each of these parameters is independent of the other element choices (e.g.
the best colour does not depend on the shape being chosen), then the performance landscape
has a single global peak. In modelling terms, this means that there are no interdependencies
and K is therefore 0. Even an extremely myopic search process, which never looks beyond
points in the immediate proximity of the current point, eventually finds this peak. Once we
find the best colour, for instance, we can hold that constant; find the best shape and so on,
sampling at most 12 points in this systematic manner.

However, if the best choice on one element depends on what has been chosen for the others –
the best colour actually does depend on the shape, which in turn depends on engine size and
so on, then the performance landscape will now have many points that are local peaks; they
are better than all their neighbouring points but not necessarily the best in the design space
overall. The K parameter regulates the degree of interdependencies and the higher the K, the
more local peaks and the more “rugged” is the design space. One will now need to sample
many more points and there may be no guarantee of finding the best point, unless all 81
points are sampled. Worse, a myopic search process that only looks at points in the proximity
of the current point (local search) will often get “stuck” on a local peak. The key difference
between the two situations lies in the existence of slopes; in the first case the slopes point
infallibly to the global peak (no interdependence, K = 0), but not in the second case
(interdependence, K > 0).

The NK model has found wide applicability in organizational research, because many
managerial problems such as new product development, process innovation and
organizational routines, strategy formulation, or the design of organizational forms, activity systems, and HRM systems appear to share a common structure: In all of these tasks, managers must make choices over many elements that jointly and interdependently contribute to overall performance, much as the design features in our car example do. In the following, we outline how the NK model depicts the task environment, the agent’s representation of the task environment, the choice process, and the transformation of representations and choices over time.

1. Task Environment:

The NK model depicts a complex combinatorial task (Kauffman & Weinberger, 1989). The N parameter captures the number of performance-relevant task elements. For example, N could represent product features, resources and capabilities, or distinct elements of a business policy. A conventional simplifying assumption is that these elements are usually binary – they are either present (1) or absent (0). To pick up on the car example above, the task of designing a car could be described as choosing the car body configuration (hatchback or sedan), the colour (black or white), and the engine type (internal combustion or electric motor). Each possible action in this task environment is a configuration or combination of these elements and the entire space of actions then contains $2^N$ combinations with distinct performance consequences. For example, with $N = 3$, the entire task environments consists of 8 distinct combinations, each associated with a unique binary string (e.g., [101], representing a sedan (1) with black colour (0) and an electric motor (1)). Combinations represent alternative product designs differing in features. Product [111] and product [000] offer radically different features, while product [110] is more similar to the first product, differing in only a single element. To ease exposition we will refer to a combination of choices such as [111] as a “policy” alternative, or in brief a policy. A policy thus represents an action in the task environment, and can be thought of as the equivalent to an arm in a multi-armed bandit model.
The NK model offers considerable flexibility so that task environments with varying levels of interdependence between its elements can be represented. This is achieved through a computational algorithm for the construction of the performance functions or “fitness landscapes”; the mapping of policy alternatives onto performance values that accounts for interdependencies among its elements. This provides the researcher with a simple way to vary the complexity of the task environment (Simon, 1962) or problem difficulty (Page, 1996).

Specifically, parameter K regulates the level of interdependencies among task elements in the payoff function. The payoff contribution of each element $e_i$ in a policy is affected by both the state of the element itself and the states of K other elements. Payoff contributions for elements are drawn from a (usually uniform) probability distribution and the performance of the entire policy alternative is then simply the average of all payoff contributions. For example, with K = 1, payoff contribution of the car body also depends on the engine type—indicating interactions between these choices – large engines mean car bodies with less passenger room, etc.

Generally, with higher values of K, there are more local peaks and performance differences among neighbouring policy alternatives, differing only in a single element, become more pronounced (Kauffman, 1993; Rivkin, 2000). This gives the landscape the property of ruggedness, with lots of peaks and surrounding valleys.

In many NK models, the interdependency structure is often randomly determined—any K other elements may influence the payoff contribution of a choice. Alternatively, the modeller may also choose to impose (by assumption) a specific structure on the task environment if the purpose of the analysis is to study the implications of these differences. For instance, the consequences of the degree of task decomposability or its hierarchical properties can be explored in this way (e.g. Ethiraj and Levinthal, 2004; Rivkin and Siggelkow, 2007; Ghemawat and Levinthal, 2008).

2. Representation:
As with any model of bounded rationality, the NK models assumes that agents do not have accurate representations of their task environment. Given the complexity and structure of the Task Environment, a crucial assumption in the application of the NK model to organizational and management problems is that an agent only constructs a limited representation of a small subset of the $2^N$ possible combinations and their associated outcomes at any point in time. Usually this subset is generated through a search process defined in terms of the neighbourhood of the status quo alternative. Thus, the agent’s representation does not include a full listing of all possible policy alternatives and their associated payoffs. For instance, the designer of the automobile is assumed to not know the payoffs associated with all possible combinations of design choices (such as black, sedan, medium engine, big trunk etc.).

In the majority of organizational NK models the agent only represents at any point in time the current policy and a randomly selected proximate policy (which differs from current policy in one element), and the associated performance outcomes of these two policies. For example, the car designer may consider a car configuration that changes the colour from white to black, while keeping other elements constant. Thus the agent’s representation at any point in time may be thought of as consisting of $\{a_1, a_{-1}\}$ and a corresponding $\{\pi_1, \pi_{-1}\}$, where these are the current policy alternative (location) and a randomly selected one-step neighbour of the current policy, and their associated payoffs (which are accurately known). The determination of the payoff of the neighbouring policy combination is often assumed to occur in an (un-modelled) “offline” process, which one may imagine features either an experiment within a “model” of the problem held by the agent, or an experiment that the agent actually undertakes but which can be reversed.

There are some important exceptions to local search. For instance, one may allow for more “distant” search or long jumps where multiple elements are changed at the same time (Cyert and March, 1963: 170; Levinthal, 1997). In models with long jumps (e.g. Levinthal, 1997),
the representations may thus (also) include a randomly selected non-proximate combination and an accurate associated payoff. For instance, the car designer may consider changing both the colour and the engine type, while keeping the car body as it is. Gavetti and Levinthal (2000) offer a depiction of agent’s representations of a landscape in which the agent has an initial coarse but unbiased representation of the landscape – they do not know the associated payoff for every combination of elements (i.e. policy alternative), but they can see aggregate payoffs for clusters of combinations of elements. To illustrate, the car designer might broadly understand that streamlined shapes and red colour seem to do better when put together, but not the various ways in which shape and colour interact with other elements (such as engine size, trunk size, weight).

Relatively, the representation could be inaccurate, as the associated payoff of the policy may not reflect their true payoffs. Knudsen and Levinthal (2007) model a search process where the associated payoffs in the representation may be subject to evaluation errors. The agent generates a local alternative to the status quo, and the evaluation of its performance is subject to an error function to capture the “screening” ability of the agent, with less able agents making more mistakes in judging the payoff of an alternative. This may lead to adoption of inferior alternatives in the search process.

More competent, less boundedly rational decision-makers might be modelled as having knowledge of more than one proximate policy and their associated outcomes in addition to the current policy in their representation. For instance, in the organizational models developed by Rivkin and Siggelkow (2003; 2005; 2006) a more competent manager has a greater number of policy alternatives in mind. This captures the degree of bounded rationality, i.e., a more competent manager develops and evaluates a greater number of policy alternatives.

In sum, Representations of agents in NK models vary in terms of scope (how many and which policy alternatives are in it at any point in time) and accuracy (whether the associated payoffs
are known and valid). However, it is important to emphasize that the notion of representation is fairly limited in the NK model, because agents do not actively learn about interactions among choice elements to guide search in the task environment, and represent only a very small portion of the task environment at any point in time.

3. Choice process

The choice process is usually assumed to follow a maximizing choice rule, selecting the policy with the best payoff in the representation. As we have noted, this representation may or may not be accurate, depending on whether evaluation errors occur (e.g. Knudsen and Levinthal, 2007).

4. Transformation procedure:

The set of alternatives in the agent’s representation is modified through a process of recombinant search (Fleming, 2001), involving the sampling of new alternative policies by recombining individual elements. The baseline assumption here is the notion of local neighbourhood search – agents seek new alternatives in close proximity to and anchored on the status quo. This is because local alternatives are easier to evaluate for boundedly rational decision-makers. More distant search which involves changing multiple policy elements at the same time may also be allowed for. The relative frequency of these types of search, and their triggers are often the key features studied in NK models.

There is learning in the NK models, because agents change their representations and hence actions as a consequence of previous actions taken and payoffs achieved. However, unlike reinforcement learning in which agents obey the Law of Effect, in NK models agents learn based on the principle of (Local)Hill Climbing - where they always search for improvements relative to their current situation, often in their immediate neighbourhood (Levinthal, 1997).

Behavioral plausibility of NK models
While individual behavior has been extensively studied in experimental work on human problem solving (c.f., Newell and Simon, 1972; Kotovsky et al., 1985; Langley et al., 2014), these investigations have generally been restricted to problems that have clearly defined goals and the agents know when these have been accomplished. In contrast, the NK model assumes that the solution – the global optimum – is not known—neither its location nor its value (c.f. Selten, 2001). This leads to two key behavioral assumptions of the NK model of individual agent adaptation - **local search** and **exploratory choice**.

Experimental studies by Busemeyer et al. (1986) and Busemeyer and Myung (1987) on resource allocation tasks with interdependencies found strong evidence for local hill-climbing behavior in both studies. Further, participants tend to select a similar direction in allocating resources if performance increased, while failure induced search in a new direction. Billinger, Stieglitz and Schumacher (2014) studied search behavior in NK task environments with various levels of complexity. They found strong evidence for the primacy of local search. Yet, they also found evidence for a form of adaptive, **problemistic search**: Individuals gradually search more distant, exploratory alternatives as a response to poor performance. These results suggest that the search for new alternatives may indeed be sensitive to performance feedback, as much of the empirical research in the organizational learning literature has assumed (Greve, 2003), but which is a notion that has not been explored to any great depth in the NK modelling literature. The studies also showed that the search process does not seem to produce a deep understanding of the overall landscape, as is implicitly assumed in the principle of local search.

Finally, the broader experimental evidence (not based on NK tasks) also supports the idea of local search. There is evidence that (1) the starting point matters because it defines the region in which people will begin searching (2) the more favourable the initial feedback is the more localized search becomes, such that the region of search becomes smaller, and finally that (3) local search reduces the region in which new experiences are sampled, and this prevents the
decision maker from reaching their optimal preference point (Ariely, Loewenstein, & Prelec, 2003; Hoeffler, Ariely, & West, 2006; Klayman and Ha, 1987).

Gavetti et al. (2012) note in their recent review of the literature that there is extensive evidence that organizational change is triggered by performance feedback and that low performance stimulates search (also see Greve, 1998; 2003). We have already noted the significant empirical literature on organizational ambidexterity, and the evidence it produces for explorative behavior at the organizational level. However, there is limited direct evidence of the local nature of organizational search and the conditions under which it becomes less local (e.g. Stuart and Podolny, 1996; Fleming and Sorenson 2004; Katila and Ahuja, 2002; Rosenkopf and Nerkar, 2001).

In sum, the key behavioral assumptions of the NK model of individual agent adaptation—namely local search and exploratory choice appear to rest on good empirical foundations. However, the experimental and empirical literature also finds to a larger extent than assumed in the models that search is adaptive and guided by performance feedback. The results here speak in favour of the notion of problemistic search as proposed by Simon (1955), Cyert and March (1963), and Greve (2003), and suggest that NK models could take on this property as a modelling feature.

**Key trade-offs and insights**

After examining behavioral plausibility we now discuss the trade-offs and insights from the NK model that we believe offer promising avenues for empirical research.

It is well known that finding the global optimum on an NK landscape is “NP hard” (Rivkin, 2000) which essentially means that it is impossible for a decision-maker to actively try out all possible configurations. For example, the problem of designing cars obviously entails many more elements than just deciding on car body, colour, and engine type. Thus organizational
scholars are most interested in the search process that unfolds in NK landscapes as well the extent to which agents in the model succeed at finding better choices. The major results from models of search in NK landscapes centre on the question of how properties of the Task Environment – in particular the ruggedness of the performance landscape – impact the effectiveness and outcomes of various search procedures. In a simple landscape with no interdependence between choices (K = 0) local search inevitably leads all agents to the global optimum. For rugged landscapes with multiple local peaks (K > 0) agents facing the same task environment are likely to end up at different local optima (which may differ in performance properties). The initial Representation (i.e. the starting location) influences where in the landscape an agent searches locally for performance improvements. This leads to strong path-dependency in the subsequent search process and the associated Transformation of the initial Representation. Because of these path-dependencies, heterogeneity in policies and in performance emerges as stable outcomes in rugged landscapes (Levinthal, 1997).

A central challenge for boundedly rational agents in rugged NK landscapes is therefore to prevent getting stuck on a low-performing local optimum and to increase the likelihood of identifying the global optimum. The models in the literature have identified two broad conditions to achieve this. First, an agent must be able to broaden search beyond the immediate neighbourhood of an existing alternative (e.g., develop more elaborate representations that go beyond the neighbourhood, in order to avoid trapping on local peaks). Second, the agent needs to be able to stabilize search around a better-performing alternative in choice selection. Readers will recognize that in essence, this is a version of the exploration-exploitation trade-off that was introduced in our discussion of bandit models. In both the NK and multi-armed bandit, exploration refers to a process of seeking alternatives that are not indicated by the current representation, though the manner in which the task environment and its representation are modelled are of course very different.
Figure 3 offers a stylized representation of this trade-off in an NK model. Early in the search process, local search rapidly increases performance from the starting position, because the task environment typically offers some refinement potential. However, over time, local search exhausts refinement potential as agents get stuck on local peaks. Further performance improvements based on local search are then no longer possible. In contrast, an adaptive search rule that triggers gradually more distant search as a response to failure and reverts back to local search after success (Billinger et al. 2014) improves performance more slowly, but tends to outperform in the long run as distant search helps agents to move off low-performing local peaks.

![Figure 3: Performance of local and adaptive search in complex task environments](image)

The modelling literature has explored several individual and organizational attributes that may help to stimulate adaptive search. Indeed, it may be fair to say that the philosophy behind most of the models in this tradition has been to take particular organizational features (or their consequences) and translate them into behaviorally plausible search processes that break free of local hill climbing, such as analogical reasoning – in which solutions that have worked in one domain are transplanted to another (Gavetti and Levinthal, 2000; Gavetti, Rivkin and Levinthal, 2005); divide and conquer heuristics in which a problem is broken down into smaller sub-problems, on the assumption that each can be solved in isolation (Ethiraj and
Levinthal, 2004); and means-end analysis, in which one works backwards from the desired
result to current best action (Denrell, Fang and Levinthal, 2004).

For instance, Gavetti and Levinthal (2000) show that simple, low dimensional
Representations – limited knowledge about the rugged performance landscape – help agents
to identify more attractive starting positions for further local search. In Rivkin and
Siggelkow’s (2003) model, less boundedly rational managers have better representations of
the task environment in the neighbourhood of the current policy. Knudsen and Levinthal
(2007) propose an approach in which agents evaluate a new policy in the neighbourhood with
error and thereby sometimes adopt a lower-performing action. This allows them to move off
low-performing peaks. This is beneficial in rugged landscapes. However, errors in the
selection of actions should not be too large, since agents may then be incapable of holding on
to superior choices.

As with the multi-armed bandit model, the insights from the NK model for the designers of
learning systems, including organizations can be inferred from this central trade-off between
local and non-local search. When entering poorly understood complex task environments, the
parameters of the search process determine success. In particular, avoiding being trapped by
local peaks with low payoffs, while also avoiding falling off peaks with good payoffs, is
critical. The models suggest that organizational processes that support analogical reasoning,
the use of divide and conquer heuristics, and sampling practices in surrounding environments,
produce this balance and therefore offer advantages in the search process. These ideas can and
should be empirically tested.

We turn next to models of mutual adjustment.

2. Models of Mutual adjustment
Models of mutual adjustment between agents capture important organizational phenomena such as inter-unit or inter-agent collaboration, socialization and knowledge exchange, imitation, and delegation and control. In fact, any organizational process in which multiple agents influence each other’s adaptation processes would qualify as a candidate to be analysed using models of mutual adjustment. For instance, managers of different product divisions within a company seeking to realize synergies by coordinating their marketing campaigns; teams of engineers developing different sub-systems; new employees adjusting to an organization’s culture and vice versa. Organizational models of mutual adjustment either explicitly or implicitly build on models of individual adaptation. Many recent models of mutual adjustment in organizations have relied on either NK or multi-armed bandit models to capture individual agent behavior, but also using some form of mutual adjustment assumptions between multiple agents.

In each of these models, two or more interacting agents jointly face either a bandit or an NK task; therefore the discussion of Representations, Choice and Transformation processes from the family of individual agent adaptation models carries over exactly here (and we do not repeat those). The key difference lies in the nature of the Task Environment that each agent faces, which now includes the other agent(s)- possibly without the focal agent being aware of this.

The discussion of behavioral plausibility of mutual adjustment models is also largely the same as in individual adaptation models, with the important exception of the task environment, which as noted above now includes the other agent(s). The social psychological evidence shows that we treat risky interactions with nature differently than we do interactions with other people, implying that social risk is different from non-social risk (Bohnet and Zeckerhauser, 2004; Blount, 1995). However, without exception, the mutual adaptation models that we review below assume that the agents act exactly as they would in a task environment in which they were alone.
There are two justifications for this. First from a modelling point of view, it is plausible that the overall learning or search process of an agent is the same regardless of whether their interaction is with Nature or another agent, but the parameters of the search process may change. For instance, agents may explore less, learn more slowly or search more locally when they know there are other agents mutually adjusting to them, but the basic processes of local search and reinforcement learning would not change. Second, one may interpret the model as pertaining to situations of such task environment complexity as to render the agent effectively incapable of meaningfully recognizing the existence of other agents (for instance if there are a large number of other agents one cannot easily communicate with). As Simon (1990) famously remarked, the task environment and the agent’s internal processes work jointly like the blades of a scissor to produce observable behavior. Above certain levels of environmental complexity, the agents’ behavior, whatever the internal processes, may look the same as it does in simpler contexts.

Besides the inclusion of other agents, there are other secondary variations in the task environments (namely interdependence, communication and imitation possibilities) of these models such that it is useful to discuss the key trade-offs and results from models of coupled learning/search, social learning, and imitation separately within this model family.

i. Key insights from coupled learning/ search models

One important class of mutual adjustment models, which we call “coupled learning/search models”, assumes that agents are engaged in a search for combinations of good interdependent choices, but do so under communication constraints, arising from specialization, spatial distance, concerns about alignment of interests, etc. Communication constraints are a critical element of these models. Absent such constraints, coordination of search is a trivial problem.
On practical grounds these models of mutual adjustment are equivalent to models of individual adaptation in a task environment that just so happens to include another agent (but the focal agent typically doesn’t really “know” this). The feedback each agent sees in any given time period is thus partly dependent on the other agent’s (unobserved) actions, resulting in the possibility of learning inappropriate lessons from feedback. Thus, the challenge for the agents is to coordinate their search processes, in a multi-period situation with feedback in every period, given interdependence and communication constraints.

Lounamaa and March (1987) presented an early model that explicitly examines mutual adjustment processes in search and learning, and pointed to the hazards of rapid learning (similar to that discussed in the section on bandit models of individual adaptation). Lave and March (1975) also reported the hazards of rapid learning in matching coordination games and offer similar explanations (i.e., superstitious learning- in which agents learn inappropriately from false negatives and false positives.

Puranam and Swamy (Forthcoming) used a coupled multi-armed bandit model (i.e., the payoff to each agent from selecting an arm also depends on what arm the other agent has selected) to show that even incorrect initial Representations of the Task environment held by both agents may be better than representations that do not discriminate at all between the alternative choices. This is because the latter are more likely to generate “false negatives”- suppose the first agent takes the correct action accidentally, but if the second agent does not, the feedback on joint actions indicates a failure, which causes the first agent to also discard their own correct choice. This effect grows stronger as the learning rates of agent’s increase (as the agents learn more rapidly from feedback that includes false negatives), and weaker as the communication between the agents improves (so that they recognize a false negative as such). Knudsen and Srikanth (2014) developed this line of argument further by showing that coupled learning processes needs to balance efforts at reducing mutual confusion by aligning Representation of the Task Environment (and thereby restricting exploration) and at reducing
joint myopia by slower learning or reduced knowledge transfer (and thus promoting exploration). These results are a manifestation of the trade-off reported in Figure 2 for the multi-armed bandit model.

Lee and Puranam (Forthcoming) model coupled learning in a hierarchy, in which a superior defines strategy and a subordinate implements it. The superior thus has a view on which arm to choose in a bandit task, transmits this view to a subordinate, who then selects the arm but with some probability of error (i.e. imprecise implementation). The superior then revises beliefs based on observed outcomes, without knowing the actual action chosen by the subordinate. The central insight in their model is that given this separation between belief (superior’s) and action (subordinate’s), effective implementation by the subordinate has benefits beyond the well-known effect of enabling exploitation of good strategies. Curtailing exploration by the subordinate paradoxically enables the discovery of better strategies by the organization, as it allows more effective learning from feedback on the value of current strategies. It therefore makes sense, the authors argue, to insist on precise implementation of strategies in organizational hierarchies even when their validity is unknown.

Rivkin and Siggelkow have studied coupled search processes on rugged landscapes using the NK model, with agents engaged in local search (2003; 2005; 2006). They find that mechanisms that cause interdependent agents to take each other’s payoffs into account (such as broad, inter-unit incentives) prevent one agent from taking actions that harm the performance of the other; but interestingly the absence of such instruments can help prevent premature settlement of the dyad onto a local peak. Siggelkow and Rivkin (2009) extended the analysis to a particular form of a hierarchical coupled learning problem, where higher
level choices partially constrain lower level choices, to show that such a coupled search process can obscure the real impact of the higher level choices. Thus the impact of choices about the first layer of organizational structure (i.e. divisional vs. functional) may be obscured by adaptation at the departmental level. For example, the performance over time of an organization depends not only on the higher level structure (e.g. divisional vs functional) but also on the manner in which lower level units adapt to their own task environments. It cannot therefore be used purely as a signal of the effectiveness of the former.

In summary, the key insights from this set of models are that in situations of interdependence, learning confronts certain unique challenges as well as opportunities. As a consequence, certain features of the mutual adjustment process are expected to play a key role in shaping its success, such as the rates of learning, the nature of initial beliefs, and the hierarchical differences between the agents mutually adjusting to each other. These conjectures can be tested empirically.

ii. Key insights from social learning models

A second class of mutual adjustment models allow for communication between independent agents, but focus instead on the rates at which agents learn from each other – in particular how they may copy or share their representations of the task environment. The actions of the agents themselves are not interdependent in these “social learning” models (Bandura, 1977), instead each agents payoff depends on their own action, but since all agents are in the same environment, they can learn from the successful actions of other agents. The classic model by March (1991) on exploration and exploitation is of this form: agents essentially learn from superior, better-performing agents in their Task environment, by copying their actions.

The models of social learning also yield the learning trade-off obtained in the coupled learning literature, i.e., moderate rates of individual learning lead to better representations of reality. The reasoning is the same, which is that rapid learning can give rise to superstitious
learning (while very slow learning may not take advantage of learning opportunities at all).

This commonality is not surprising given that the key building block for both kinds of models is an individual level reinforcement learning process. Subsequent work in this tradition has deepened insights around these social learning processes. Miller et al. (2006) and Lazer and Friedman (2007) added a social network perspective to the underlying mechanism, with individuals only learning from each other if they are socially connected. Fang, Lee and Schilling (2010) and Schilling and Fang (2013) developed the social network perspective further by showing how subgroup isolation and information distortion enhance exploration in organizational learning. Isolated groups with few direct ties between them preserve individual variations in representations longer. Isolated groups thereby slow down the learning process, leading to more correct representation of the task environment. Likewise, Information distortion generates variation as agents learn from each other and has a similar effect.

In summary, who one is connected to, and at what rate one learns from others as opposed to from one’s own experiences are factors that should shape the effectiveness of learning among a community of learners.

iii. Key insights from Imitation/Competitive Learning models

Finally, several models of mutual adjustments have considered a particular form of social learning, which occurs when the interests are not aligned between teacher and learner and indeed they may be competing- e.g. the case when laggards in an industry try to imitate the leaders. Using the NK model as a framework, the central insight here is that complexity may serve as an effective barrier to imitation (Rivkin, 2001; Lenox et al., 2007). Specifically, the finding is that partial or imperfect imitation of complex practices (perfect imitation is unlikely by definition when complexity is high) may not necessarily improve performance (Rivkin, 2000).

Csaszar and Siggelkow (2010), on the other hand, demonstrate some value to imitation, albeit in low complexity environments. They do this by pinpointing two distinct mechanisms
influencing the value of imitation. First, imitators benefit directly from copying superior practices, especially in low-complexity task environments. Second, imitation even of inferior practices is also an instrument for dislodging agents from inferior local peaks and thereby promotes more distant search. Relatedly, Posen et al. (2013) develop the intriguing insight that imperfect imitation may be better than perfect imitation in some situations. Specifically, imperfect imitation allows the imitator to preserve and build upon unique and valuable practices currently not implemented by superior firms.

A central element of adaptive rationality is the presence of aspirations that determine when search and imitation stops. Aspirations of an agent may not just be informed by the individual search and learning process –their own historical aspirations (Lant, 1992; Greve, 1998) – but they may also be conditioned socially on the performance of others (Festinger, 1954; Greve, 1998). When aspirations are socially formed, they induce a unique mutual adjustment dynamic, since agents that uncover superior alternatives raise the reference point for the entire ecology of learners (Levitt and March, 1988; Lant and Mezias, 1992). This can, in turn, reignite search and learning by others, thereby creating a self-reinforcing dynamic of mutually adjusting aspirations and increasing performance. Thus, competition serves an important role in spurring innovation and imitation, not only because it raises the expected returns to search (as assumed in traditional rational choice models), but also because the success of others raises the aspiration of the focal firm. This is crucial in a task environment in which no one may know the upper bound on what is truly achievable (i.e. the height of the global peak is unknown).

These results point to possibilities for empirical examination of the effects of complexity on competition and innovation.

3. Information Aggregation
Like mutual adaptation models, in information aggregation models agents are part of each other’s task environments. As a family, these models provide a general framework to understand how individual evaluations are aggregated into organizational level outcomes (Knudsen & Levinthal 2007; Christensen & Knudsen 2010; Csaszar 2013; Csaszar and Eggers, 2013). In the earlier versions of these models, while the agents interact with each other, they do not necessarily adapt to each other— their Representations and Choice processes typically do not undergo any change over time. Recent papers assume a transformation process for changing the agent’s representations and choice processes, qualifying this family of models for inclusion in our review of adaptive rationality models.

Individual evaluations of options in these models are opinions about the absolute viability or relative attractiveness of alternatives. (If an individual’s evaluation is the final one and not subject to appeal, we may call it a decision by the organization). For instance, when individual managers in an organization evaluate projects, investment alternatives, or candidates to hire, they are generating evaluations (which may or may not be the final decision of the organization, depending on the distribution of decision rights, e.g., is the relevant manager the CEO?).

The models we discuss in this section study the impact of different ways in which individual evaluations are aggregated into an organizational level decision. These different decision architectures are of course a very important part of organizational structure, and so these models may be seen as formalizations of the consequences of different organizational structures to aggregate evaluations. The models draw on the seminal work of Sah and Stiglitz (1985, 1988) and depict decision-makers as holding fallible beliefs about the value of an alternative. The emphasis is therefore clearly on the evaluation, rather than the generation of alternatives (Knudsen and Levinthal, 2007).
As with all models of bounded rationality, the representations the agents hold (in this case, about what makes an alternative worthy of positive evaluation) may be flawed. In general, fallible decision-makers could make two types of errors: They may reject a superior alternative (Type I error) or accept an inferior alternative (Type II error). An everyday example may help to convey the intuition. Imagine a group of employees who are screening credit card applications. Each makes an individual evaluation—should an application be accepted or rejected? Each employee individually is fallible, and may make either a Type I or a Type II error. But how does the structure of the group influence its aggregate performance? In other words, should we get each application seen independently by an employee who makes a final decision on it, or should we get an application seen by multiple employees who must all agree before it is accepted? These different decision architectures will have very different aggregate Type I (reject superior alternative) and Type II (accept inferior alternative) error rates.

A decision architecture consists of a set of rules that influence information flows and decision rights in an organization. For example, in a three-member linear hierarchy, a new project is first evaluated by a lower-level manager. Only if she accepts the project, it gets passed on to a middle manager. Otherwise, the project is rejected and not evaluated further. The middle manager, in turn, then also decides to reject it or accept it to send it on to the senior manager for final approval. In contrast, in a decentralized polyarchy, all three managers would evaluate the project and acceptance by one is sufficient for project approval. These building blocks describe a wide range of possible organizational architectures, with many hybrid forms such as committee voting falling between the extreme forms of the hierarchy (all agents must accept and pass on to the next) and the polyarchy (any agent may accept) (Knudsen and Levinthal, 2007; Christensen and Knudsen, 2010; Csaszar, 2013; Csaszar & Eggers, 2013).

Task Environment:
The task environment that each agent faces is one in which alternatives arise typically sequentially rather than simultaneously, and in which there are no interdependencies between alternatives. Each agent faces a stream of projects with some associated payoff value. Evaluations involve characterizing these into good and bad projects. For example, in considering investment projects, a project could be evaluated as worthy of acceptance if it offers a positive net present value.

**Representation and Choice process:**

In these models, the representation and choice process are combined into a single modelling element called a screening function. Decision-makers may differ in their ability to represent accurately the relationship between attributes of the project and its value, denoting differences in skills, experience, or talent. This differential ability is usually modelled as a screening function. This is a function that gives us the probability of a project being accepted as a function of some underlying characteristic. For instance the characteristics could be quality, and agents may differ in their ability to correctly assess this. An ideal screener is one who accepts the project with probability=1 if it exceeds the quality threshold, else rejects it. Fallible screeners, however, will accept even below threshold quality projects with some probability, and reject above threshold quality projects with some probability.

**Transformation process:**

In many models, the screening function of an agent remains constant over time. Christensen and Knudsen (2013) introduced a model in which the screening ability of agents develops as a function of the decision architecture they are in. Agents are assumed to learn from the alternatives that they evaluate. The assumption here is that alternatives represent learning opportunities and that more experience translates into a higher screening ability.

**Behavioral plausibility**
Models of information aggregation make few psychological assumptions besides the idea that individual evaluations are often imperfect because of bounded rationality, but may improve with experience. This assumption appears to have strong empirical support (c.f., Kahneman and Klein, 2009). The central behavioral idea here is that agents identify and learn about relationships between objectively identifiable project cues and the associated value of incoming projects. For example, the default risk of a credit application by a company is associated with certain cues (such as current profitability, growth, investments etc.). The behavioral research shows that learning leads to improved screening ability when the relationship between cues and project value are stable, when feedback is rapid and unequivocal (i.e., less noisy), and when opportunities exist for prolonged practice with a similar set of projects (Osman, 2010; 2014; Newell and Shanks, 2014).

**Key trade-offs and insights**

The models provide many insights into how different aggregation structures convert individual evaluations into organizational decisions.

Figure 4 offers a stylized depiction of the core result from this class of models. Acceptance probability, here, is a function of the project value: a perfect evaluator would reject all projects with a value below zero and accept all projects with a value above zero. A single agent with imperfect screening ability is depicted as a straight line with a positive slope, indicating that there is a positive probability of rejecting (accepting) good (bad) projects.
In a hierarchy, every agent has to accept an alternative for the entire organization to adopt it. This lowers the probability of accepting bad projects and minimizes Type II error, but this way of aggregating beliefs comes at the costs of a higher rate of Type I errors. In contrast, the acceptance of an alternative by a single agent is all that is required for its adoption in a polyarchy. This results in a higher rate of Type II errors and less Type I errors. Put differently, the polyarchy in the aggregate is more explorative and less risk-averse than the hierarchy (Csaszar, 2012; 2013).

Csaszar and Eggers (2013) extended the information aggregation perspective by also considering how delegation, majority voting, and the averaging of opinions impact information aggregation and the decision-making quality. They show that delegation is the most effective structure when there is diversity of expertise, when accurate delegation is possible (i.e., when it is known who has the expertise), and when there is a good fit between the firm’s knowledge and the knowledge required by the environment. In other words, trusting the expert is a good option under the very restrictive conditions identified above. Otherwise, voting or averaging may be the most effective structure. This result formalizes and specifies the notion of wisdom of crowds (Surowiecki, 2005).
Christensen and Knudsen’s (2013) model with transforming screening functions allows agents to learn from experience and get better at screening. The basic model structure is further enriched by allowing agents not only to accept or reject an alternative, but also to pass on an alternative when they do not feel confident about their own judgment. These advances therefore mix problems of mutual adjustment and information aggregation to get a richer understanding of organizational decision-making over time. A key insight is that if an agent only sees projects already screened by others based on his current location in the decision architecture, her screening function may transform over time to one which would be inappropriate if she were placed elsewhere in the structure.

4. Linkages between Empirical Analysis and Models

In this section, we conclude by summarizing and discussing the linkages between empirical analysis and models. These occur in two directions. First, there is the linkage between modelling assumptions and the evidence on actual adaptive behavior by individuals and organizations. Second, there is the linkage between model generated results and testable empirical implications.

i. The behavioral plausibility of modelling assumptions

For each model family, we examined the structure of the models in terms of four common elements, the Task Environment, Representations, Choice processes, and Transformation (of Representation) processes - with the goal of understanding to what extent the assumptions in these models are consistent with the evidence about how individuals and organizations behave. Our review paid close attention to the key behavioral assumptions in these models of adaptive rationality, and these may be summarized as follows:

a) **Complexity of the Task Environment** to a point where the agent’s Representations are necessarily incomplete/incorrect; as a consequence agents may not represent the
entire search space, or possess knowledge about the payoffs associated with actions, or be aware of the existence and actions of other agents that shape their own Task Environments, or see the true value of alternatives. As we have noted, this assumption has more to do with the translation of an empirical context appropriately into a model; certainly there are organizational contexts where these assumptions are highly plausible, and others where they are less so.

The remaining assumptions pertain to individual psychology or organizational behavior relating to the nature of adaptive rationality:

b) *Exploration in choice*: Choice processes within the Representation stray from pure exploitation (i.e. maximization or greedy action selection) and feature a degree of exploration.

c) *Problemistic search* to modify representations – search and exploration is triggered by performance falling below aspiration levels.

d) *Local hill climbing to modify representation*: the discovery of new alternatives occurs primarily through the principle of local improvements; (a relaxation of this tendency and the willingness to engage in distant search may be interpreted as a form of exploration).

e) *Reinforcement learning to modify representations*: the discovery of the associated values for known alternatives occurs through the Law of Effect - alternatives with better than expected outcomes are more likely to be tried again, those with worse than expected outcomes are less likely to be tried again.

The good news for researchers in the organizational sciences, modelers and otherwise, is that the evidence for these assumptions at both the individual and organizational level is in general supportive. We have documented the empirical evidence, both experimental and non-experimental, that gives credibility to these assumptions in the relevant sections where we discussed each model.
To be sure, there are places where the evidentiary base is stronger than others, and much replication is necessary before these assumptions can be safely granted the status of axioms. For instance, individual level evidence on problemistic search, and organizational level evidence of hill-climbing behavior both can be strengthened. It is still largely a mystery as to when both individuals and organizations display broadly similar (though certainly not identical) adaptive search and learning behavior, and when they do not since it is apparent that the strong unitary actor (e.g. the heroic CEO with fully centralized decision powers) is such a rarity.

The aggregation processes within organizations that produce broad similarity across scales still need to be carefully investigated, and is a fascinating research agenda in itself. As we will note below, scholars interested in this particular research question can draw on work done in social psychology and organizational behavior that has compared group versus individual decisions.

But in answer to the first question that motivated this review—namely are the assumptions that underlie models of bounded rationality in organizations behaviorally plausible?—we must answer, with cautious optimism, in the affirmative. The behavioral plausibility of modelling assumptions could serve as an encouragement for researchers to put in effort at further refining tests of behavioral regularities so that modellers can strengthen and make more precise the assumptions in their models. Likewise, empirical researchers in organization science, we suggest, should be receptive to at least engaging with these models as sources of testable predictions.

---

8 We mean similarity beyond that induced by the language of description. For instance, if we want to compare individual and organizational screening functions, we must look for similarities beyond those generic to any screening function.
ii. The empirical implications of model results

If organizational models of (adaptive) bounded rationality fare well in terms of their use of behaviorally plausible assumptions, we should be optimistic that their predictions find empirical support. Yet surprisingly, attempts to empirically test the predictions of these models have been rather limited. In part this may be because of the subtlety involved in interpreting the results of such a test, even if conducted. To elaborate on this, we consider two kinds of tests.

First, when we test the implications drawn from, say a multi-armed bandit or an NK model of individual adaptation, we are testing the predictions resulting from the interaction between a Task Environment with certain assumed properties, and a set of behavioral rules (captured in the nature of Representations, Choice processes and Transformation process). These are of course the famous twin blades of Simon’s scissors- the environment and the agent’s internal processes jointly produce observable behavior (Simon, 1956; 1990). However since the latter (behavioral rules) were drawn from an existing evidentiary base, the test is first of all about the former (i.e. assumptions made about the task environment), and whether these correspond to the empirical context being used to test the model predictions.

Further, one may also test if variations in the parameters of the search or learning process produce expected results when coupled with particular properties of the task environment.

Laboratory experiments, where one can carefully construct environments, appear to be a good method to test model predictions. To be useful and correspond to real world problems, these laboratory based studies could be combined with carefully constructed case studies (c.f., Edmondson & McManus, 2007 about the utility of mixed methods in fields that are neither very mature nor very new).

Second, when the Task Environment includes other agents (in models of mutual adjustment and information aggregation), then our tests of model predictions are implicitly testing if the
environments are indeed such as to cause agents in them to be unable to distinguish other
agents immersed in the same environment from the environment itself. Put simply, we are
testing the assumption that the agents behave non-strategically with respect to each other in
the task environment of interest, since that is what the behavioral rules assume in the models
of mutual adjustment and information aggregation. The limited data that examines differences
in agent behavior when interacting with human actors compared to non-humans (e.g., games
where human agents play with a machine or with nature) suggests that this assumption is not
valid in all task environments (e.g., Blount, 1995; Bohnet & Zeckerhauser, 2004).
Confronting the predictions of these models with such empirical data may lead to a
recognition that the behavioral rules used by individuals in simple task environments that they
recognize as containing other agents (e.g. Blume, Duffy, Franco, 2009) may be different from
more complex situations or in which the other agents fade, as it were, into the task
environment (Osman and Ananiadis-Basias, 2013). The empirical literature is somewhat
limited in that the task environment faced by decision makers has hitherto not been very
complex. It is not clear that the observed difference will generalize to very complex
environments. Thus, how much iterative reasoning about other’s intentions is both plausible
(and tractable) will then be a key issue for modelers and empiricists to consider together, by
varying task environments systematically.

With these observations in mind, we outline the broad contours of the kinds of empirical tests
one could conduct (and are indeed being conducted) of the results from the existing models.
We outline the correspondence between model parameters and phenomena in naturally
occurring data, and these are also summarized in Table 1.

----- INSERT TABLE 1 ABOUT HERE -----
For models of individual adaptation based on multi-armed bandits, the key properties of the task environment are noise and changes in feedback conditions. The key behavioral parameters of the learning process are the extent of exploration and the rate of learning. Empirical tests could thus, for example, involve comparing differences in performance between adaptive agents (individuals or organizations) in task environments with varying rates of changes or degrees of noise in feedback (e.g. industries or investment opportunities with varying levels of environmental change or noise in performance feedback) as a function of the agent’s tendencies to make choices other than the past successful ones (i.e. exploratory choices) and responsiveness to feedback (i.e. learning rates).

The key property of the task environment in NK models is the interdependency parameter $K$, which creates ruggedness in the performance landscape. The search parameters of interest are those that affect the balance between making exploratory “leaps” on the landscape instead of local hill climbing. Thus, tests of these models would require us to exploit variations in task environments in terms of their interdependence structure (e.g., highly decomposable task environments, such as the production of software, vs. less decomposable ones such as the development of an advertising campaign), along with variations in properties of the decision-making agent that would lead to more or less adherence to local hill climbing behavior.

A few studies have already been conducted that offer support for the predictions from the NK model. Using patent data, Fleming and Sorenson (2001) found that high levels of technical interdependencies reduce the usefulness of inventive efforts (i.e., tracking model results that show that exploitation is not a useful strategy in complex environments). In a related study, Fleming and Sorenson (2004) showed that investments in basic science are particularly valuable in technological fields with high interdependencies. In such task environments, local search quickly traps inventors on local peaks. Science provides a cognitive representation (Gavetti and Levinthal, 2000) of the technological search space and thereby helps inventors to identify more attractive regions in the landscape. Lenox, Rockart, and Lewin (2010) used
survey data to directly test their own model that enriches the basic NK setup with oligopolistic competition among organizations (Lenox, Rockart, Lewin, 2006). They found that average profitability is highest with moderate interdependency, and that the dispersion and skewness of profits is especially pronounced with high levels of interdependencies. These contributions provide empirical support for NK model predictions about how complexity affects managerial and organizational behavior and performance outcomes.

Models of mutual adjustment through coupled learning based on the multiple-armed bandit or the NK framework create an organization comprising multiple agents. The search process of this organization depends on the variance in the parameters of the following features of its internal structure: the nature of reward interdependence between the agents, their relative rates of learning or exploration, degree of heterogeneity of its networks, and the distribution of decision rights between them. These structural features of organizations are observable. By jointly examining these attributes as well as variations in the noise and interdependence in task environments, the model predictions can be empirically tested. Social learning models, through their increasing use of network structure as a key feature of the models, also point to an obvious empirical testing strategy: to link network structure to innovation and knowledge heterogeneity across agents (Fang et al., 2010). Models of imitative/competitive learning make the structure of the task environment a key feature of the models, and so lend themselves to empirical testing by examining the heterogeneity of agent’s stable sets of choices as a function of the interdependence in their task environments (Csaszar and Siggelkow, 2010).

In models of information aggregation, the key aspects of the task environment pertain to the rate of arrival and value of projects to be evaluated. Organization design theories have built extensive typologies of environmental conditions such as turbulence and munificence that capture such variations (e.g., Aldrich, 1979; Dess and Beard, 1984; McCarthy et al. 2010). These typologies and the instruments used to measure variance on these dimensions can be
used to empirically verify model predictions about how rate of arrival of projects and the
value of projects to be evaluated lead to particular decisions of organizations. Also, the
aggregation structures that combine individual evaluations (the key element of these models)
are related to the distribution of decision rights in organizations-as visible in authority
hierarchies and committees. Measuring differences in authority hierarchies and committees
and combining them with differences in environmental munificence and turbulence can be
and indeed has been one method used to test predictions derived from information
aggregation models (c.f., Csaszar, 2012; Reitzig and Maciejovsky, 2014).

An alternative to looking for naturally occurring data is to set up experiments in the
behavioral laboratory\(^9\) to test predictions. As we noted at the outset of the review, there is a
natural complementarity between modeling and experimental work in that they both deal
with interaction structures among a small number of agents, in carefully controlled task
environments (Billinger et al, 2014; Reitzig and Maciejovsky, 2014). Experiments thus
provide a very direct test of the predictions emerging from models, as indeed has been
discovered to their great advantage by researchers in behavioral game theory (Camerer,
2003).

To make our idea about complementarity concrete, we consider some possible
experimental paradigms that may be useful starting points (if not destinations) on such a
journey.

As we have outlined, the m-armed bandit and the NK landscape are two basic model
families that have been used to understand adaptation by unitary decision makers.

\(^9\) Field experiments would be even stronger if feasible to set up in a way to allow for precise tests of
mechanisms (e.g. Bloom et al. 2013).
Fortunately experimental tasks exist that correspond quite closely to the task environments of these models. The IOWA gambling task (Bechara et al., 1997) is one such example.

Bandit tasks and the IOWA gambling task share similar structural properties. Both present decision-makers with a fixed number of choice alternatives, and in both each choice alternative have a fixed rate of reward which is unknown to the decision maker. From trial to trial the decision-maker receives information (outcome feedback and/or reward information) from their choice between the alternatives, and their job is to reliably select sequentially from the alternatives so that they maximize their rewards. The appeal of bandit tasks, and by extension variants of the IOWA gambling task, is that they lend themselves very well to examining both psychological phenomena (such as risk taking, impulsivity, and exploration behavior) and modelling including a determination of what is optimal behavior in different choice environments. This laboratory task has many features which can be carefully manipulated (e.g., payoff structure, feedback, framing, training length, cost of information search, and time pressures) that allow researchers to understand the types of strategies that individuals develop while exploring and then exploiting their environment.

As outlined above, the NK model has emerged as a primary modelling approach to study complex management problems such as new product development (e.g. Almirall and Casadesus-Masanell, 2010; Claussen et al., 2014), organization design (e.g. Rivkin and Siggelkow, 2003), and strategy making (e.g. Gavetti et al., 2005). It may also be a useful experimental platform to study search, learning, and decision-making in complex task environments. Busemeyer et al. (1986) used a simple resource allocation task in which participants have to divide up a fixed number of resources among distinct alternatives (such as allocating time to research and teaching). This task can be made more or less complex by allowing for payoff interactions among the alternatives. The resource
allocation task thereby shares some common features with the NK model, because experimental researchers can vary the number of alternatives (equivalent to the N parameter in NK models) and the presence of interaction (K parameter) and can be used to study predictions from NK models. Billinger et al. (2014) proposed an experimental product design task building directly on the NK model. Participants, in their experimental setup, must design a product by (re-) combining various attributes and then receive performance feedback about specific combinations. The task allows for the easy manipulation of, for example, the number of design elements (the N parameter in the NK model), interactions among them (the K parameter), different reward structures, or the information available to participants (e.g. prior knowledge about the task environment such as specific interactions among elements or the payoff of the global optimum, information about the designs and current performance of other participants etc.). These experimental tasks may allow for the detailed study of how human decision-makers grapple with complex problems in various settings such as new product development, process improvements, or the discovery and modification of new organization designs or business models. Likewise, getting a better understanding how individuals and groups search, learn, and choose in complex situations might lead to a refinement of existing formal models and theories.

Experimental tests for the aggregation models can take their inspiration from research that uses games to investigate how groups differ from individuals as decision making units. There are many elegant demonstrations of differences between groups and individuals using classical economic games such as the prisoner’s dilemma, dictator, ultimatum, trust, centipede and principal-agent games (Kugler et al, 2012; Wildschut et al, 2003). In the prisoner’s dilemma game (PDG) two players choose between a cooperative and a non-cooperative option, and the combination of choices both players make have different payoffs. In adapted versions a unitary decision to cooperate or defect is made by a group, which can be compared
with decisions made by individuals. Wildschut et al (2003) and others since have shown that
groups are more competitive than individuals in the PDG. Ways of amplifying and
attenuating this discontinuity effect have been connected to key factors associated with group
dynamics (e.g., strong conflicts of interest, distributed responsibility of decisions taken,
relaxed attitudes to violations of norms of fairness, level of communication). Moreover,
when looking across studies comparing groups with individuals in economics tasks the
emerging pattern seems to be that groups make unitary decisions that are closer to rational in
game theoretical terms (Kugler et al, 2012; Maciejovsky et al., 2013). Using tasks such as the
PDG to look at aggregation issues is extremely valuable, because it is a simple starting point
from which to examine different aggregation structures, such as consensus, voting, or
hierarchy in order to compare how groups using these rules act, and if there are any
predictable differences to how individuals do so. Another approach is to use the group versus
individual method used in the PDG with experiments that use paradigms developed to test
behaviors in NK environments. Some emerging research in this area examines how groups
differ from individuals on local versus distant search. Using the experimental NK task,
Billinger et al. (2015) examined how two-person teams search for performance improvements
in rugged landscapes. They find that teams search less locally and perform better than
individuals in identical tasks.

A second, and as yet largely uncharted area of exploration is to construct task
environments that allow the study of mutual adjustment processes. Experimental tasks in
which pairs (or more numbers) of agents take interdependent actions and adjust to
feedback are of course staple in social psychology and behavioral game theory (e.g., Roth
and Murnighan, 1978; Murnighan and Roth, 1983); however tasks that map closely to
bandit or NK models are yet to be constructed. In sum, we can think of the space of
opportunities here along these two dimensions: 1) Individual vs. multi-agent decision
making entity and 2) Adjustment to environment vs. adjustment to another decision
making entity (i.e., mutual adjustment). The quadrant that involves groups interacting with
groups is sparse at this point, but is perhaps the closest to inter-organizational phenomena
of interest such as competition and collaboration.

A challenge that organizational scientists may face in relying on experiments is the skepticism
that readers and reviewers may display about whether interactions between small numbers of
agents (often two) may have anything to say about “real” organizations, as opposed to say
“merely” teams, dyads or groups. (This is in addition to the usual concerns about external
validity that are justifiably flagged with any experimental work). A thorough consideration of
the issue is more than we can provide in this paper, but we will note the following: a review
of the most influential definitions of organization reveals certain commonly occurring criteria
for an entity to be considered an organization, such as the existence of goals whose attainment
depends on the efforts of multiple agents, but the number of agents (beyond the “more than
one” requirement) is never among these criteria (Puranam, Alexy, Reitzig, 2014). In our view,
there is no basis (besides convention) on which one can say that a three-person firm is an
organization but a four-person team is not; to the extent these are goal directed multi-agent
systems, they are both organizations, albeit solving their problems of division of labor and
integration of effort in different ways (also see Raveendran, Puranam and Warglien,
Forthcoming).

There are of course important and completely legitimate questions to be addressed on whether
the manner in which a three-person organization works is the same as an n-person
organization (scaling), or whether the behavior of the system changes when each agent is
itself an organization (recursion). However, we believe we can make a start at answering
them by taking the first steps into the lab, possibly guided by intuitions derived from models.
The prospect for fruitful collaborations between experimental researchers and modelers in
advancing organization science seems bright.
5. Conclusions

We began the paper by suggesting that the literature on adaptively rational models of organizations based on Simon’s (1916-2001) pioneering work on bounded rationality was deserving of critical examination on account of its central claim that it is based on ‘correct’ behavioral assumptions. We also suggested that the extensive research in this tradition has generated many precise insights that could be empirically tested. To facilitate this examination and provide organizational scholars with a tractable typology of models we classified and examined model families. Specifically, we reviewed three important families of models of adaptive rationality in organizations—individual adaptation, mutual adaptation, and information aggregation. These do not exhaust the model families prevalent in organizations research, but represent conceptually important and influential ones. And finally, we discussed opportunities for research that strengthen the links between formal modelling in organizations research, and its behavioral foundations.

Our review suggests that the adaptively rational models are based on empirically validated behavioral assumptions in the main. We have identified a few assumptions that bear further testing. We have also identified the key insights and predictions generated by different models and provided information about the empirical tests (albeit limited) of these predictions. Finally, we discussed opportunities for future research and outlined how empirical tests can be constructed to take advantage of these opportunities.

Our hope is that this review will facilitate greater dialogue between modelers and empiricists and that organizational theories and empirical studies will attain greater precision and provide better predictions as a result.
References

(Note: Contributions are coded for model families for easy reference: BA = Multi-armed Bandit Model, NK = NK model; MA = Mutual Adjustment Model; IA = Information Aggregation model)


Fang, C. (2012). Organizational learning as credit assignment: A model and two experiments. Organization Science, 23(6), 1717-1732. (BA)


Lee, E., Puranam, P. (Forthcoming). The implementation imperative: Why one should implement even imperfect strategies perfectly, Strategic Management Journal. (BA, MA)


Raveendran, M., Puranam, P., & Warglien, M (Forthcoming), Object salience in division of labor: experimental evidence, Management Science


Table 1: Summary of model families

<table>
<thead>
<tr>
<th>Model</th>
<th>Task environment</th>
<th>Representation of task environment</th>
<th>Choice process</th>
<th>Transformation</th>
<th>Key trade-offs</th>
<th>Key insights</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Models of individual adaptation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bandit model</td>
<td>Actions are discrete alternatives with uncertain and unknown outcomes</td>
<td>Subjective beliefs about outcomes (e.g. discriminative between choices, and bias of representation)</td>
<td>Probabilistic choice based on representation</td>
<td>Reinforcement learning to update representation</td>
<td>Exploitation of current representation versus exploration to update representation</td>
<td>Value of exploration (Sutton &amp; Barto, 1998)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Key parameter: Propensity to explore by taking actions other than those currently believed to be best (e.g. deviation from past successful actions, novelty seeking)</td>
<td>Learning without being misled by positive and negative feedback</td>
<td>Dangers of learning too rapidly from noisy feedback (Denrell, 2005; Posen &amp; Levinthal, 2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Key parameter: Responsiveness to feedback (e.g. incentives for success and failure)</td>
<td></td>
<td>Hot stove effect (Denrell &amp; March, 2001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Risk aversion as an outcome of learning (March, 1996)</td>
</tr>
<tr>
<td>NK model</td>
<td>Actions are complex combinatorial alternatives with certain but unknown outcomes</td>
<td>Limited representation of available alternatives and of structure of task environment</td>
<td>Maximizing choice with potentially imperfect evaluation</td>
<td>Recombinant search to create and test new alternatives</td>
<td>Local versus distant search</td>
<td>Value of combining local and distant search (Levinthal, 1997; Siggelkow &amp; Levinthal, 2003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Key parameter: Quality of initial representation (e.g. coarseness, discrimination among choices, and bias of representation)</td>
<td></td>
<td>Key parameter: Propensity to change many elements at a time (e.g. uncoordinated search with specialists vs. coordinated search, political deadlocks leading to local search)</td>
<td>Perfect versus imperfect alternative evaluation</td>
<td>Efficacy of cognitive search (Gavetti &amp; Levinthal, 2000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Key parameter: Interdependence between choice elements (e.g. decomposability of design space or value chain)</td>
<td></td>
<td></td>
<td></td>
<td>Value of imperfect evaluation for exploration (Knudsen &amp; Levinthal, 2007)</td>
</tr>
<tr>
<td>Models of mutual adjustment</td>
<td>Outcomes of alternatives depends on the choices of other agents</td>
<td>Agents take actions that are interdependent, adapt to feedback</td>
<td>Slow versus rapid learning from own choices in updating representations</td>
<td>Value of slow learning (Lounamaa &amp; March, 1987)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------------------------------------------------------</td>
<td>---------------------------------------------------------------</td>
<td>----------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coupled learning/search models</td>
<td>Draws on models of individual adaptation</td>
<td>Key (additional) parameters: Reward interdependence between agents, relative rates of learning, initial representations, communication constraints, distribution of decision rights</td>
<td>Coordination versus exploration</td>
<td>Benefits of incorrect representations (Puranam &amp; Swamy, 2014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social learning</td>
<td>Presence of other agents in a collaborative setting</td>
<td>Key (additional) parameters: Connectivity to other agents, relative rates of learning</td>
<td>Slow versus rapid learning from choices of other agents</td>
<td>Value of weak ties (Miller et al., 2006; Fang et al. 2009) and information distortion (Schilling &amp; Fang, 2012) to preserve slow learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imitation/Competitive Learning</td>
<td>Presence of other agents in a competitive setting</td>
<td>Agents update representations by observing the choices of others in a competitive setting</td>
<td>Imitation versus innovation</td>
<td>Complexity as imitation barrier (Rivkin, 2000; 2001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Key (additional) parameters: Observability, imitation breadth</td>
<td></td>
<td>Imitation to reignite exploration (Csaszar &amp; Siggelkow, 2010)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Value of imperfect imitation (Posen et al., 2013)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Models of information aggregation

<table>
<thead>
<tr>
<th>Decision structures</th>
<th>Actions correspond to projects to be selected</th>
<th>Decision-makers differ in ability to evaluate outcomes of alternatives</th>
<th>Maximizing choice with imperfect evaluation</th>
<th>Updating based on reinforcement learning</th>
<th>Type I (rejection of superior alternative) versus Type II error (acceptance of inferior alternative)</th>
<th>Design of reliable decision structures (Christensen &amp; Knudsen, 2010)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Key parameters: Rate of flow and outcomes of alternatives</td>
<td>Key parameters: Experience and expertise in alternative evaluation</td>
<td>Key parameters: Experience and expertise in alternative evaluation (note representation and choice are combined in these models)</td>
<td>Key parameter: Responsiveness to feedback (e.g. incentives for performance)</td>
<td>Type I (rejection of superior alternative) versus Type II error (acceptance of inferior alternative)</td>
<td>Design of reliable decision structures (Christensen &amp; Knudsen, 2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Decision structures and exploration (Knudsen &amp; Levinthal, 2007; Csaszar, 2012)</td>
<td>Limits to majority voting and averaging of opinions (Csaszar &amp; Eggers, 2012)</td>
</tr>
</tbody>
</table>